DOCUMENT RESUME

ED 101 017 95 TM 004 569

AUTHOR	Ozenne, Dan G.; And Others
TITLE	Achievement Test Restandardization: Emergency School Aid Act (ESAA) National Evaluation.
INSTITUTION	System Development Corp., Santa Monica, Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
REPORT NO	TM-5236-006-00
PUB DATE	1 Nov 74
CONTRACT	OEC-0-73-0831; OEC-0-73-6336
NOTE	136p.
EDRS PRICE	MF-$0.75 HC-$6.60 PLUS POSTAGE
DESCRIPTORS	*Achievement Tests; Background; Content Analysis; Cultural Differences; Culture Free Tests; Elementary Education; *Federal Programs; Item Analysis; Mathematics; *Minority Group Children; Norms; Pretesting; Questionnaires; Reading; Sampling; Standardized Tests; Statistical Analysis; *Test Bias; Testing Problems; *Test Selection
IDENTIFIERS	California Achievement Test; *Emergency School Aid Act; ESAA

ABSTRACT

To meet several pre-evaluation goals of the national Emergency School Aid Act (ESAA) Evaluation study, the following activities that were undertaken are described: selection of an achievement measure; a pretest of this measure to assess student needs in schools eligible to receive funds under the Act (schools with a minority enrollment of 50 percent or more); an important research effort directed toward possible ethnic and/or cultural bias in the measures; and the establishment of test norms to aid in the interpretation of student and school performance relative to the appropriate sub-population. California Achievement Test (CAT-70) Levels 2 and 3 subtests measuring Reading Comprehension, Mathematics Computations, and Mathematics Concepts along with a questionnaire describing student background were administered to a nationally representative sample of ESAA-eligible students. A descriptive analysis of the resulting data is provided and the research conducted to investigate possible bias in the selected measures is documented. Appropriate measures of the debiased scales and the implications of their use are discussed. Finally, methodology employed for scaling the tests for use in interpreting performance within the ESAA-eligible population is presented. (RC)

achievement test restandardization

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED101017

TM 004 569

TM

emergency school aid act national evaluation

BEST COPY AVAILABLE

SYSTEM DEVELOPMENT CORPORATION · 2500 COLORADO AVENUE · SANTA MONICA, CALIFORNIA 90406

ED101017

EMERGENCY SCHOOL AID ACT (ESAA) NATIONAL EVALUATION

ACHIEVEMENT TEST RESTANDARDIZATION

(TM-5236/006/00)

Dan G. Ozenne

Nancy C. Van Gelder

Alan J. Cohen

NOVEMBER 1, 1974

UNITED STATES OFFICE OF EDUCATION CONTRACT NUMBERS

OEC-0-73-0831 AND

OEC-0-73-6336

TABLE OF CONTENTS

		Page
	Overview of the Emergency School Aid Act National Evaluation	vi
	Executive Summary.	x
I.	INTRODUCTION	1
II.	ACHIEVEMENT TEST SELECTION	3
III.	SAMPLE SELECTION AND TEST ADMINISTRATION	8
	A. SAMPLE SELECTION	8
	B. TEST ADMINISTRATION.	20
IV.	DESCRIPTIVE ANALYSIS	22
	A. STUDENT SAMPLE	22
	B. STUDENT BACKGROUND CHARACTERISTICS	25
	C. STUDENT ACHIEVEMENT LEVELS	34
	D. SUMMARY OF STUDENT ACHIEVEMENT LEVELS AND BACKGROUND CHARACTERISTICS	37
V.	DEVELOPMENT OF UNBIASED ESAA ACHIEVEMENT MEASURES	47
	A. RATIONALE AND METHODOLOGY.	47
	B. STATISTICAL ANALYSIS	49
	C. CONTENT ANALYSIS	57
	D. RESULTS.	60
VI.	PREPARATION OF NORMS	64
VII.	SUMMARY	66
	REFERENCES	68

TABLE OF CONTENTS (continued)

		Page
APPENDIX A.	LETTERS TO SUPERINTENDENTS OF THE DISTRICTS, CHIEF SCHOOL OFFICERS, AND COMMISSIONERS OF THE HEW REGIONS	69
APPENDIX B.	RESPONSES TO STUDENT BACKGROUND QUESTIONNAIRE.	75
APPENDIX C.	RESPONSES TO STUDENT BACKGROUND QUESTIONNAIRE BY ETHNIC GROUP.	82
APPENDIX D.	ITEMS IDENTIFIED AS POTENTIALLY BIASED, WITH REVIEWERS' COMMENTS	85
APPENDIX E.	SCORE-TO-PERCENTILE-RANK CONVERSION TABLES AND SCHOOL-MEANS-TO-PERCENTILE-RANK CONVERSION TABLES FOR GRADES 3, 4, AND 5 . . .	93

LIST OF FIGURES

		Page
Figure 1.	Example of Graphical Detection of Item-by-Group Interaction . . .	50

LIST OF TABLES

Table 1.	Students in U.S., in Restandardization Universe, and in Sample	11
Table 2.	Districts in U.S., in Restandardization Universe, and in Sample	11
Table 3.	Summary of Strata of Schools for Restandardization Sample	15
Table 4.	Number of Districts and Schools in the Final Restandardization Sample	19
Table 5.	Achievement Test and Background Questionnaire Respondents	22
Table 6.	Comparison of Means for Total Sample and Subsample Responding to Background Questionnaire	24
Table 7.	Significant Differences Between Groups on Student Background Questionnaire Items	26
Table 8.	Percentages for the Modified Ethnic Categories.	29
Table 9.	Order of Frequency of Materials in the Home by Grade Level. . . .	31
Table 10.	Order of Frequency of Materials in the Home by Ethnic Group . . .	33
Table 11.	Summary of Raw Score Test Characteristics for Restandardization Sample	35
Table 12.	Range and Skewness of Achievement Test Scores by Grade Level. . .	36
Table 13.	Achievement Levels of Restandardization Sample Relative to National Norms.	38
Table 14.	Achievement Test Performance by Ethnic Self-Identification. . . .	40
Table 15.	Ethnic Group and First SES Measure.	44
Table 16.	Ethnic Group and Second SES Measure	45
Table 17.	Analysis of Covariance for Fourth-Grade Total Reading Scores. . .	46
Table 18.	Examples of Aberrant Item Characteristics	52

LIST OF TABLES (continued)

		Page
Table 19.	Summary of Number of Items Identified as Suspicious, by Level and Subtest	53
Table 20.	Items Identified as Potentially Biased and Reasons for Identification.	54
Table 21.	Summary of Test Characteristics	61
Table 22.	Test Score Means for Full and Derived Scales for Two Reference Groups	62

Acknowledgements

The work described in this report was performed in conjunction with the Emergency School Aid Act (ESAA) Pilot Program Evaluation under the general direction of Dr. John E. Coulson, Project Director, System Development Corporation (SDC). This work was supported under U.S. Office of Education contract OEC-0-73-0831, with Dr. Michael J. Wargo serving as Program Officer. The authors wish to thank Drs. Coulson and Wargo for their support and assistance during this effort.

Additionally, the authors wish to acknowledge the assistance provided by other SDC staff. Dr. William Kent, assisted by special sampling consultant Earl Houseman, was responsible for selecting the restandardization sample and supplying documentation of the sampling paradigm. Dr. Dean Lee was responsible for conducting the content review phase of the item bias research.

The views reflected in this document are not necessarily the views of the U.S. Office of Education.

OVERVIEW OF THE EMERGENCY SCHOOL AID ACT (ESAA) NATIONAL EVALUATION

The Emergency School Aid Act (ESAA) was enacted into law in June of 1972 to provide elementary and secondary school districts with financial assistance to: (1) meet the special needs incident to the elimination of minority group segregation and discrimination, (2) encourage the voluntary reduction, elimination, or prevention of minority group isolation, and (3) aid children in overcoming the educational disadvantages of minority group isolation [P.L. 92-318, Sec. 702(b)]. While the Act as amended in 1974 (P.L. 93-380 Sec. 641) authorizes the appropriation of one billion dollars for fiscal year 1973 and a similar amount for the period ending June 30, 1976, actual appropriations have amounted to 270 million and 234 million dollars for fiscal years 1973 and 1974 respectively, with the fiscal year 1975 appropriation pending. Since funds are annually appropriated for obligation and expenditure during the fiscal year succeeding the year of appropriation, the major thrust of the Act began during school year 1973-74 and is expected to continue through school year 1976-77.

Seventy-four percent of the Act's annual appropriation is reserved for two subprograms, the Basic Grants (59%) and the Pilot Programs (15%). The Basic Grant program is essentially a desegregation program designed to reduce minority group isolation, meet the needs incident to the elimination of segregation and discrimination, and to aid school children in overcoming the educational disadvantages of minority group isolation. In contrast, the Pilot program is a compensatory education program designed to improve the academic achievement of children in minority isolated schools (i.e., schools with over 50% minority enrollment).

The sums annually appropriated pursuant to the Act are apportioned to States on the basis of the ratio of their number of minority group school aged children to the number of such children in all States. Local school districts compete for the funds apportioned to their State through grant applications to their

HEW Regional Office. In applying for an ESAA grant a local school district must demonstrate that it has needs related to the Act's objectives and that it has designed a program based upon authorized activities that shows promise in achieving one or more of the Act's objectives.

Evaluation Objectives

The Act authorizes a national evaluation of its programs which is supported by an annual one-percent reservation of appropriated ESAA funds. As designed by the U.S. Office of Education (USOE) and conducted by the System Development Corporation (SDC), the national evaluation focuses on an integrated evaluation of the ESAA Basic and Pilot programs and has the following general objectives:

- Determination of the short and long term national impact of the program in terms of the Act's objectives, namely, reduction of minority group isolation, elimination of discrimination, and improvement of basic skills in elementary and secondary schools.
- Identification and description of the needs of students in or from minority isolated schools; the characteristics of local programs, including their resource allocation's relationship to needs; and the interrelationships of those factors with program impact.
- Documentation and dissemination of information relating to unusually successful local programs and program components that appear to be related to success.
- Determination of the relative effectiveness of three forms of educational intervention--desegregation, compensatory education, and their combination--as compared to no special intervention in minority isolated schools.
- Investigation of the relationships among regular school expenditures, supplementary ESAA expenditures, and program impact in an attempt to determine local program cost/effectiveness and the minimum supplemental expenditures necessary to ensure some measure of program success.

In an attempt to achieve those objectives, data are being collected from a nationally representative sample of ESAA-funded school districts over a period of two to three school years.

Evaluation Methods and Procedures

Data on achievement, school climate and discrimination, and reduction in minority group isolation, have been collected annually since school year 1973-74 in a nationally representative sample of approximately 75 Basic and 42 Pilot elementary schools and 54 Basic secondary schools in 85 ESAA-funded school districts. Within each school in the evaluation, samples of approximately 60 students in each of grades 3, 4, and 5 or 10, 11, and 12 were randomly selected across sections within grade to participate in the evaluation. Students are followed longitudinally through those grade bands, with grade 5 and 12 students leaving the sample each year. In any one year there are approximately 27,000 students, 4,000 teachers, 172 principals, and 85 local ESAA coordinators, district business managers, and superintendents in the evaluation sample.

The selection procedures for schools within districts consisted of classifying all ESAA-eligible schools in terms of estimated prior student achievement, estimated socio-economic status of enrolled students, and percent and type of minority composition. Pairs of matched schools (schools similar in the above dimensions) were then randomly selected and within each pair, one school was randomly assigned to the treatment (ESAA funding) condition and the other school to the control (no ESAA funding) condition. These procedures resulted in a true experimental design with comparable treatment and control schools in each district.

At the beginning and end of each school year, mathematics and reading achievement tests and questionnaires are administered to all students in the evaluation sample. Monthly, a student attendance and exposure log is completed for or by each student in the sample to obtain data on the types of activities students are exposed to, and the frequency and duration of exposure to each activity. Near the end of each school year, a battery of questionnaires is administered to superintendents, district business managers, local ESAA coordinators, principals, teachers, and students in the sample. Those questionnaires provide data on district, school, and classroom minority group isolation, program operation, resource allocation, and student and staff background characteristics.

Data analysis will focus on the major objectives of the study through use of instruments tailored to measure the Act's three major purposes. Annual analysis will include comparison of pre-post change in outcome measures among treatment and control schools, comparison of relative effectiveness of different intervention approaches, identification of unusually successful local projects, and determination of the relationships between program characteristics and program impact. In addition to annual analyses, cumulative impact of the program will be determined on the basis of the longitudinal data collected from the sample. Finally, cost-effectiveness of the program at the local and national level will be determined annually.

Approximately seven months after post-test data collection each year, System Development Corporation in conjunction with USOE will produce evaluation reports summarizing ESAA impact. Each report succeeding the first will address the subjects of cumulative impact and comparative impact after successive years of program implementation.

Evaluation Design Features

The ESAA evaluation design has a combination of features that make it an advance in the state-of-the-art in national evaluation. Previous national evaluations have included one or more of the design features of the ESAA study, but no other study to date has integrated all of the following highly recommended evaluation procedures: a sample representative of the population affected by the program; annual pre-post data collection on impact measures; longitudinal data collection; randomly selected schools and random assignment of treatment and control conditions; three measures of impact directly related to the program's national objectives; an achievement test restandardization that resulted in a supplementary set of norms for minority isolated schools which will be used in conjunction with existing national norms; use of an achievement test specially modified to reduce its possible bias against minority students; and finally, a combination of classical and Bayesian data analyses techniques. It is expected that the particular combination of design features that constitute the national evaluation of ESAA will result in less ambiguous results than previous national evaluations and a firmer basis upon which Congress and Administration can judge the ultimate effectiveness of the Act.

EMERGENCY SCHOOL AID ACT ACHIEVEMENT TEST RESTANDARDIZATION

EXECUTIVE SUMMARY

Respected members of the education, test development, evaluation, and minority communities have at various times charged that existing standardized achievement tests are inappropriate for the assessment of minority student academic performance. In general, this charge is based upon the fact that minority group students are often under-represented during two important phases of the test development process, namely, test item selection and test standardization. As a consequence of such minority group under-representation, standardized achievement test items are said to be biased against minority students and test norms are said to be inappropriate for minority students and for schools with high minority student enrollments. In short, these critics claim that existing standardized tests are developed by, with, and for white middle-class America.

A counter argument to the bias criticism, made by an apparently equal number of qualified individuals, is that even granting that minority groups are under-represented in most test development efforts, standardized achievement tests have an important function in school systems regardless of their minority concentration. Such tests provide a standard, albeit a middle-class white American one, by which students and schools across the nation can be compared to each other. According to this argument, achievement tests are a valid criterion for assessing the ability of all students to achieve in our society.

Recognizing the apparent validity of both arguments, and realizing that debate has yet to resolve the issue, it was decided early in the planning stages of the Emergency School Aid Act (ESAA) national evaluation to develop an achievement test that would satisfy the interests of both camps. The primary objective of the activity was to select the best existing reading and mathematics achievement test battery currently available for evaluation of the ESAA program and then to improve the sensitivity, reliability, and validity of the battery for the evaluation population--students in or from minority-isolated schools.[1]

---

[1] The Emergency School Aid Act (ESAA) defines minority-isolated schools as schools with a minority enrollment of 50% or more.

The major product of the activity would be a restandardized achievement test with (a) norms for students in the nation's schools in general, (b) supplementary norms for minority-isolated schools and students in such schools, and (c) two scoring systems, the original scoring system and one which would be less biased against minority students. If that product could be achieved, then it would be possible to assess the impact of the ESAA program using both scoring systems referenced to both sets of norms. That is, a student's score or school's mean score, original and debiased, could be compared to the norm for schools and students in general and to supplementary norms for minority-isolated schools and students enrolled in such schools.

The restandardization process consisted of several steps, the most important of which were test selection, test administration within a nationally representative sample of minority-isolated schools, identification of items biased against minority students and removal of such items from one of the test scoring systems, and development of a set of supplementary norms for minority-isolated schools and children in such schools. This technical report, the first major product of the ESAA evaluation, discusses the procedures employed during re-standardization and the results of the effort. The following paragraphs will briefly summarize those activities and discuss the limitations and potential usefulness of the product.

The test selection phase of restandardization consisted of a review of all existing standardized achievement tests and the selection of a pool of reading and mathematics subtests that appeared to be most appropriate for the ESAA evaluation. Criteria used in the initial screening process included the following: test appropriateness in terms of minority group representation in the item selection and standardization phases of test development; extent of apparent minority-group bias; relevance, interest, and meaningfulness to minority-group students; grade level and content relevance; administration time burden; and reliability, validity, and normed technical excellence. The pool of subtests remaining after initial screening on the basis of the selection criteria was then reviewed by an independent panel of test development experts.

The panel, in conjunction with the USOE and the evaluation contractor, System Development Corporation, finally selected the California Achievement Test, 1970 edition (CAT-70), Levels 2 and 3. The reading comprehension, vocabulary, mathematics concepts, and computations subtests of the battery were selected for restandardization. It should be noted that of the tests reviewed, none met any one or all of the selection criteria fully, and that practical considerations resulted in the selection of the CAT over a few other tests that ranked as well as the CAT.

Although the CAT was considered among the best tests on the basis of the selection criteria, it, like all other existing standardized tests, suffered from the fact that it was developed with, and standardized on, a sample of students significantly different from the ESAA evaluation sample. The CAT, like most other standardized tests, was designed for national use; therefore, minority groups were represented in its item selection and standardization sample in approximately the same proportion as the proportion of minority children in the nation's schools. The ESAA evaluation sample, however, was expected to be composed of well over 50% minority group students. Consequently, it was necessary to re-standardize the selected test battery on a nationally representative sample of students and schools similar to those that would eventually be selected for the ESAA evaluation sample, i.e., minority-isolated schools and students enrolled in such schools.

The selected CAT subtests were administered to a random sample of 30 students in each of grades 3, 4, and 5 in a nationally representative sample of 100 minority-isolated schools, near the end of the 1972-73 school year. Approximately 9,000 students were tested. Data so collected served as the basis for the identification of biased items and the development of achievement norms for minority-isolated schools and students in those schools.

Administration of the CAT prior to ESAA program implementation provided information for a national needs assessment of children in minority-isolated schools. These data indicated that students--minority and majority group members--in minority-isolated schools have a significant need for remedial reading and mathematics

programs. The mean achievement in reading and mathematics of students in minority-isolated schools wavered about the 20th percentile relative to existing national norms. Approximately 80% of elementary students in the nation achieve at a higher level than the students in the restandardization sample. That fact provides a clear and unambiguous indication of the national need for programs such as ESAA, which are targeted at improving the basic skills of children in or from minority-isolated schools.

The restandardization data were subjected to statistical analysis in an attempt to identify test items that might be biased against minority students. A test item was considered biased if it did not measure what it purported to measure for both minority and majority group students. After statistical analysis, suspicious items were reviewed for their content by a special panel of minority-group experts from the fields of test development and education. The panel was requested to review all items that statistical analysis suggested might be biased and to reach a consensus on those items that panel members judged to be truly biased against minority students. The items so identified were then removed to form a special, supplementary, less biased test scoring system.

Data collected during the restandardization was also subjected to standard statistical analyses and scaling procedures which resulted in the development of achievement norms for minority-isolated schools and students enrolled in such schools. Those norms will be used in conjunction with both scoring systems and with existing national norms to determine the local and national impact of the ESAA program during school years 1974, 1975, and perhaps 1976.

In evaluating the general significance and usefulness of the data reported herein, the reader should be aware of several salient limitations imposed on the restandardization by time constraints and the overall ESAA evaluation design. It should be noted that the restandardized test was designed to assess only the reading and mathematics achievement of students in or from minority-isolated schools at grades 3, 4, and 5. The restandardized test is therefore inappropriate for other student populations, grades, or subject matter areas. Further, since an existing test was debiased and restandardized, rather than an entirely new

test being developed, the items in the test may not be the best possible items for use with the ESAA subpopulation. Even the items left in the debiased test were selected by the original test developer on the basis of tryout-testing on a sample of students in which minority groups were under-represented. Nevertheless, those items included in the test that were later identified as being biased against minority students were eliminated from the restandardized test scoring system. If practical constraints had not limited this effort, an entirely new test would have been developed and item selection could have been based upon item testing with a group of students more representative of the user population, i.e., students in or from minority-isolated schools. The impact of that constraint cannot be fully known unless such an ideal test construction effort is undertaken and scores on the restandardized CAT are compared to those of the new test.

In sum, as part of the national evaluation of the ESAA program an existing standardized achievement test was restandardized and debiased on the basis of data collected from students enrolled in a nationally representative sample of minority-isolated schools. Although there were practical constraints on the completion of the restandardization, the results as described in this report will be considered an advance in test development. Judgments as to the extent of the advance and the general usefulness of the restandardization are left to the reader. Nevertheless, it is encouraging to note that on July 5, 1974, the National Association for the Advancement of Colored People (NAACP) at its 65th Annual Convention in New Orleans passed a resolution demanding a moratorium on administration of any standardized test unless steps were taken to (a) include a representative number of minority students in the standardization sample, and (b) correct the test for biases against minority-group children. Such steps have been taken in the restandardization described in the following pages.

Michael J. Wargo, Ph.D.
U.S. Office of Education
ESAA Evaluation Program Officer

I. INTRODUCTION

In response to the United States Office of Education (USOE) specifications for the Emergency School Aid Act (ESAA) Pilot Program Evaluation, restandardization testing activities were initiated in Spring 1973. These activities were directed toward the selection of an achievement test to be used in the national evaluation, a pre-evaluation administration of this instrument to a sample of ESAA-eligible students in minority-isolated schools, and the analysis of data obtained.

The restandardization testing was designed to meet several goals. The first of these was to assess the academic needs of students in ESAA-eligible minority-isolated schools, prior to program implementation. The funding of ESAA Pilot programs was predicated on the belief that minority isolation adversely affects student achievement. The restandardization testing was intended to assess the impact of minority isolation on reading and mathematics achievement. Additionally, the needs assessment would establish national baseline achievement data for students in ESAA-eligible minority-isolated schools. These data could then provide a basis for studying changes in achievement patterns after program implementation.

The second goal of the restandardization testing was to evaluate the adequacy of the achievement measures for purposes of the evaluation. Standardized achievement measures have often been accused of bias against minorities, since these measures are typically developed for the majority population. To investigate the characteristics of the instrument when used in a high-minority-enrollment subpopulation, a research effort directed at the issue of item bias was initiated. The purpose of this investigation was to determine whether there was evidence of bias in the measure. If such bias was found in a subset of the items, the biased items would not be used in scoring the measure for many purposes of the evaluation. The resulting scales would be more appropriate and sensitive measures of program impact.

A third goal, closely linked to the first two, was to provide a set of norms for expressing the scores of students in ESAA-eligible minority-isolated schools and for expressing school means. These norms would enable us to relate a student's or school's achievement level to that of other students or schools with similar characteristics. Such comparative norms would give a more appropriate baseline for noting the relative positions of the treatment and control schools in the evaluation, and would provide an interpretive scale for use with the debiased measures.

This document discusses the collection and analysis of the restandardization data. Sections II and III describe the selection of the achievement instruments used and the selection of the restandardization sample. Sections IV, V, and VI address the goals of the restandardization testing. Section IV provides a descriptive analysis of the data resulting from the administration of the achievement test and the attendant student background questionnaire. This analysis describes the sample actually obtained for testing and documents the assessed achievement levels, thereby providing both the needs assessment and the baseline data.

Section V documents the research conducted to investigate the existence of possible bias in the measures selected. The rationale and resulting methodology for detecting item bias are described. Empirical data relevant to the detection of biased items are provided, as well as the comments related to the actual items identified as biased. The appropriate uses of the debiased scales, as well as the implications of their use, are discussed.

Finally, Section VI presents the methodology employed for scaling the tests for use in interpreting performance within the ESAA-eligible minority-isolated subpopulation. Score-to-percentile-rank conversion tables are provided.

II. ACHIEVEMENT TEST SELECTION

The first step in the restandardization testing activity was the selection of the achievement measures to be used. The selection of the instruments was subcontracted to UCLA's Center for the Study of Evaluation (CSE), because of their extensive experience in the field of test evaluation, and was guided by criteria specified by USOE. In order to implement a workable schedule for test selection, examinee selection, and test administration within the limited time and resources available between contract award and the end of the 1972-73 school year, initial efforts were concentrated on selecting measures appropriate to the elementary grade levels targeted for the evaluation. Secondary-level measures were selected later. The following pages describe the process by which the measures were selected.

The selection process began with a careful review of the criteria set forth for instrument selection. The following USOE criteria were to be used for test selection:

- The selected achievement test battery should cover grades two through twelve (2-12).*
- The selected battery must have reading (comprehension and vocabulary) and mathematics (concepts and computations) subtests for all grade levels.
- The subtests should be independent of specific curricula; i.e., they should be basic skill tests.
- The subtest levels should have some grade-level overlap to minimize the possibility of floor and ceiling effects.
- Total administration time for the combined reading and mathematics tests should be approximately two (2) hours or less.

*Only elementary-level tests were considered in the initial CSE ratings. Selection of secondary-level tests was a separate task.

- The selected subtests should have minimum ethnic-group bias.
- Minority groups should be represented in the restandardization sample.
- The selected subtests should have acceptable test reliability and validity.
- Norms and scales of the battery should be adequate for the proposed study.
- Subtest content should be relevant, interesting, and meaningful to today's minority-group students.
- The subtests should be easily administered, scored, and processed.

A multi-stage strategy was adopted for selecting the elementary-level achievement tests. First, a master list of potentially applicable tests was compiled. This list was obtained primarily from CSE Elementary School Test Evaluations (Hoepfner, Strickland, Stangel, Jahsen, Patalino, 1970), one of the products of CSE's Evaluation Technologies Program. To produce this book, CSE had amassed a file of all published tests appropriate for the elementary school level. After reaching agreement on criteria for judging the quality of tests, CSE had rated over 2,500 instruments on these criteria and had published the results in the book. As part of the rating procedure, tests had been placed into categories corresponding to 145 goals of elementary education. The goals were intended to represent all possible student outcome goals at the elementary school level. Several goal areas closely match the outcome dimensions relevant to the ESAA evaluations. The test list was assembled from the following CSE goal areas:

CSE Goal	ESAA Outcome Dimensions
Arithmetic Concepts	Mathematics Concepts
Arithmetic Operations Operations with Integers	Mathematics Operations
Reading Comprehension	Reading Comprehension
Understanding Ideational Complexes	
Inference Making from Reading Selections	
Recognition of Word Meanings	Reading Vocabulary

In addition to the tests contained in the CSE goal areas, several of the instruments that were specifically designated by USOE but did not fall into these goal areas were added to the list. The total list consisted of 66 separate test batteries.

The initial list of 66 tests was then reduced by applying several absolute cutting criteria. If a test did not meet one of these criteria, it was immediately eliminated from further consideration by CSE. In order to stay in the list of contenders, a test had to:

- be designed for group administration,
- have alternate forms,
- be amenable to machine scoring (e.g., optical scanning),
- have percentile or grade equivalent norms, and
- be a measure of achievement rather than of intelligence.

The application of these criteria led to a reduction of the initial list from 66 tests to the following 13 instruments:

- Bobbs-Merrill Arithmetic Achievement Series
- California Achievement Test
- Comprehensive Test of Basic Skills
- Contemporary Mathematics Test
- Iowa Test of Basic Skills
- Stanford Achievement Test
- Wisconsin Contemporary Test of Elementary Mathematics
- Gray-Votaw-Rogers General Achievement Test
- Burnett Reading Series
- Sequential Tests of Educational Progress
- Nelson Reading Test

- Metropolitan Achievement Test
- Science Research Associates Achievement Series

Each of the 13 remaining tests was then rated by CSE staff members on six selection criteria representing important dimensions of test desirability. These were:

- content/construct validity
- examinee appropriateness
- alternate-forms reliability
- curricular representativeness of test
- distributional characteristics (as projected for ESAA sample)
- degree of freedom from ethnic bias

After completing its ratings of the tests on the individual criteria, CSE nominated two mathematics tests (California Achievement Test and Sequential Test of Educational Progress) and three reading tests (California Achievement Test, Comprehensive Test of Basic Skills, and Stanford Achievement Test) as best meeting the needs of the ESAA Pilot Program Evaluation at the elementary school level. A packet of materials was then sent to each member of the Test Selection Panel*; this packet included the original list of 66 candidates, the cutting criteria, the relative rating criteria, and the tentative nominations. On March 17, 1973, a meeting of the Panel was held at SDC in Santa Monica. Participants included Ralph Tyler, Robert Hess, and Charles Thomas of the Test Selection Panel, along with representatives of SDC and CSE. At this meeting, CSE staff members reviewed the entire rating procedure with the panelists and discussed reasons for certain ratings given to several of the final contenders. The Test Selection Panel concluded the meeting with a recommendation that the final selections be made from the five finalists on the basis of practical considerations such as ease and speed of obtaining the necessary quantity of instruments. (Approximately 27,000 instruments were needed.) This recommended

*The Test Selection Panel is a group of consultants retained by SDC for purposes of screening the achievement instruments. Each of the panel members was selected because of his competency and experience in test theory and his awareness of the problems associated with testing in large-scale evaluations.

course of action was adopted, and led to the selection of the California Achievement Tests in both reading and mathematics. The instruments finally selected were the CAT Level 2 and Level 3 subtests measuring Reading Comprehension, Reading Vocabulary, Mathematics Computations, and Mathematics Concepts.

Because there was considerable concern for the amount of time that test administration would take and because a student background questionnaire was to be administered in the same session, it was considered desirable to shorten the test battery. This shortening was effected by deleting certain subsections of the mathematics subtests at both levels. Specifically, the problems subsections were removed from both levels of the mathematics concepts subtests, and the subsection measuring computations involving fractions was removed from the computation subtest at Level 3. The removal of these subtests shortened the achievement battery sufficiently to allow administration within a single morning or afternoon session. It was felt that the remaining items in each subtest constituted a purer measure of the desired dependent variable.

III. SAMPLE SELECTION AND TEST ADMINISTRATION

The restandardization testing, which took place in May and June 1973, involved selecting a nationally representative sample of students in grades 3, 4, and 5 in schools with more than 50% minority enrollment, i.e., minority-isolated schools. Approximately 9,000 students in 100 schools across the United States were included in the sample. The following discussion describes how the universe to be sampled was defined and gives details on the procedures used for stratifying the universe, drawing the sample, and administering the test.

A. SAMPLE SELECTION

1. Definition of Universe

Eligibility criteria, as established by the ESAA Pilot Program, were, first, that the school have more than 50% minority-group enrollment, and second, that the school be in a district with more than 50% minority-group enrollment on a district-wide basis (or, in the case of large districts, with a minority-group enrollment of at least 15,000). Schools meeting the first criterion (more than 50% minority-group enrollment in the school) are called "minority group isolated." The term "minority group", according to the Act, means "(i) persons who are Negro, American Indian, Spanish-surnamed American, Portuguese, Oriental, Alaskan natives, and Hawaiian natives and (ii) . . . persons who are from environments in which a dominant language is other than English and who, as a result of language barriers and cultural differences, do not have an equal educational opportunity . . . ."

A rough definition of the universe sampled, therefore, is that it consisted of all students in grades 3, 4, and 5 enrolled in May 1973 in schools eligible to receive ESAA Pilot awards. However, certain refinements were added. For example, handicapped students who were not testable under the same conditions as regular students had to be excluded, in order for a standardized test battery to be used. This and other practical considerations led to the following, more exact definition:

The universe sampled included all students who in May 1973 had all of the following characteristics according to the most recent HEW Office of Civil Rights Survey*:

(1) Enrolled in regular grade 3, 4, or 5 classes or equivalent ungraded classes. (Special education classes were excluded.)

(2) Enrolled in a district in the continental United States with a total enrollment (according to last available figures) of 300 or more students. Since districts with fewer than 300 students are not systematically surveyed by the Office of Civil Rights (OCR), the reports on them are incomplete and unrepresentative. In addition, the per-student cost of testing in these districts would be very high. Only 1.2 percent of all U.S. elementary and secondary students were enrolled in 1971 in such districts. Therefore, these districts were excluded from the norms universe. Hawaii is not surveyed by OCR. Alaska was excluded for logistic reasons, as well as for its unique minority situation.

(3) Enrolled in a district whose minority enrollment (according to last available OCR figures) was either greater than 14,999 or greater than 50% of the district's total enrollment. (Note: The ESAA definition of "minority" includes several groups not specifically covered by OCR surveys, such as Portuguese, Alaskan natives, Hawaiian natives, and certain persons from non-English-speaking environments. No attempt was made to estimate the impact of this change in definition on minority enrollments as surveyed by OCR.)

(4) Enrolled in a minority-isolated school (actual condition in May 1973, whether previously reported or not). An attempt was made to use the ESAA definition of "minority" to determine current minority isolation of schools.

---

*The Office of Civil Rights Survey, conducted each school year, reports enrollment by district, and by school within district, indicating the actual numbers of students of various minority backgrounds at each grade level.

(5) Attending school and testable on the day the test was given. Absent students were not given make-up tests. Students were considered testable unless they had some handicap or severe language difficulty that interfered with their taking the test under the same conditions as regular students.

Estimates of districts, schools, and students in the universe as defined were made from the latest available OCR data. For the most part, Fall 1971 survey data were used. Preliminary 1972 data were used in a few cases to confirm or establish the universe membership of marginal districts.

A detailed analysis was made of OCR Report 71-441, which tabulates minority enrollments as of October 1971 in all districts surveyed. Within each district, all schools are listed in this report in descending order of minority percentage, and enrollments by type of minority are given. All minority-isolated schools in minority-isolated districts (or in districts with more than 14,999 total minority enrollment) were noted. For each such school, the number of students enrolled in grades 3, 4, and 5 (or in equivalent ungraded classes) was estimated. The estimate took into account the number of grade levels and the apparent proportion of 3rd, 4th, and 5th graders present in each school. Special estimating factors for schools reporting ungraded classes were based on other reported information such as grades present (e.g., an entry of "K456U" is assumed to have graded classes for kindergarten and grades 4, 5, and 6, and the ungraded equivalent of grades 1, 2, and 3). The universe so established for the continental United States totaled 1,506,751 students in 691 districts. Tables 1 and 2 indicate the propostion of the total student and district universes represented by these values.

2. Overall Design of Sample

A sampling plan was adopted for selecting 30 students at random across sections within each of the three grades in a stratified random sample of 100 minority-isolated schools, for a total of 9,000 students. This number and configuration

Table 1. Students in U.S., in Restandardization Universe, and in Sample

Student Groups	Number	Percent of Total
Students in U.S. public schools	45,428,464 (1)	100%
Students in districts with 300 pupils or more	44,886,914 (1)	98.8%
Students in 3rd, 4th, and 5th grades in U.S. public schools	11,000,000	24.2% (2)
Students in minority-isolated schools in districts in continental U.S. with 300 pupils or more eligible for ESAA Pilot Program grants	6,250,000 (3)	13.8%
UNIVERSE: Students in 3rd, 4th, and 5th grades in minority-isolated schools in districts with 300 pupils or more in continental U.S. eligible for ESAA Pilot Program grants	1,506,751 (3)	3.3%
SAMPLE: 90 students in each of 100 randomly selected schools	9,000	(0.6% of universe of students)

Table 2. Districts in U.S., in Restandardization Universe, and in Sample

Districts	Number	Percent of Total
All districts in U.S.	16,515 (1)	100%
Districts with 300 pupils or more	11,666 (1)	70.6%
UNIVERSE: Districts in continental U.S. eligible for ESAA Pilot Program grants	691 (3)	4.2%
SAMPLE: Districts containing 100 randomly selected schools	65	0.4% (9.4% of universe of districts)

Notes to Tables 1 and 2:

(1) Education Directory. 1972-73 Public School Systems, GPO, 1973.
(2) Statistics of Local Public School Systems, Fall 1969, Pupils and Staff, GPO, 1971.
(3) Estimates based on OCR Survey data.

of students was manageable within the limits of project resources, and was judged by staff sampling specialists to be adequate for restandardization purposes.

A two-stage sampling process was used in which the primary units were schools and the secondary units were students. When a school did not have all three grades of interest (3, 4, and 5), it was combined with a complementary school in the same district to form a pair of schools with all three grades. A complementary school was needed when the originally selected school had only one or two of the three target grades; the complement was always a school that had grades either immediately higher or immediately lower than the originally selected school and that either received students from or sent students to the originally selected school each year. For this discussion, "school" will refer to a primary sampling unit which is either an individual school or a pair of schools.

Some reduction in costs could have been achieved by adopting a three-stage sample, where the first-stage units would be districts, the second-stage units schools, and the third-stage units students (for example, a stratified random sample of 50 districts, each with two schools and 180 students). The prime consideration in adopting a two-stage rather than a three-stage sample was that the component of sampling error associated with variation among schools is larger than the component for variation among students within schools. Since field staff were readily available for conducting tests in widely-scattered locations, the cost and time savings from testing in fewer districts would not be sufficient to offset the lower statistical efficiency of a three-stage sample. Therefore, it was decided to reduce the between-school component of error by adopting a two-stage sample.

3. Stratification

Considerations affecting the approach to stratification included the following:

(1) As nearly as possible, every student in the defined universe should have the same chance of being in the sample. Therefore all

strata should contain approximately equal numbers of students, so that random selections in one stratum would have the same probability as random selections in all other strata.

(2) Sampling error for the two-stage design selected should be estimated from a measure of variation among schools within strata. To have an unbiased estimate of sampling error, a minimum of two schools from each stratum was necessary. On the other hand, to derive as much benefit from stratification as possible, a maximum number of strata were needed.

On the basis of these considerations, the universe was divided into 50 strata of approximately equal size (i.e., about 30,000 students each) and two schools were selected at random from each stratum, to reach the desired total of 100 schools. Schools were selected with probability proportional to size (in estimated number of students in the defined universe), following generally accepted practice for multistage sampling. This procedure was statistically efficient in terms of the goal of minimum sampling error. Also, since schools were selected with probability proportional to size and since all students had nearly the same chance of being in the sample, the number of students selected for testing was approximately the same in each sample school--a desirable feature from the standpoint of test administration.

Three kinds of information were readily available for stratifying schools: (1) geographic location, (2) ethnic composition, and to some extent (3) degree of urbanization. Information on size of school was used in the process of selection; that is, schools were selected with probability proportional to size. This was a more effective way of using school size information than using it for stratification.

Income level of the community was considered as a possible additional criterion for stratification, since it might have reduced variance slightly. However, income figures were not available in usable form. Furthermore, most schools in

the survey population were presumably in low or low-to-medium income areas, so that the income range would be somewhat limited. Also, variation in income is probably associated with the three more readily available types of information; hence potential gains from using income-level data appeared to be moderate.

The objective of stratification was to minimize variation among schools within strata. Three criteria based on the three available types of information were used: (1) percent minority, (2) degree of urbanization, and (3) geographic region. Percent minority, as a characteristic of individual schools, was broken down into five categories (90-100%, 80-89.9%, 70-79.9%, 60-69.9%, and 50-59.9%). Urbanization and geography were considered together to produce several categories: (a) metropolitan, i.e., located in a district that either exceeded 5,000 in measure of size (M.O.S.) in the defined universe or was in the same Standard Metropolitan Statistical Area as another district exceeding 5,000 in M.O.S.; (b) non-metropolitan, non-Southern (Southern States were defined as Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia); (c) non-metropolitan, Southern, medium (M.O.S. 800 to 5,000); and (d) non-metropolitan, Southern, small (M.O.S. under 800).

In addition, geographic location from West to East was used as a general criterion for grouping schools together into strata after all other criteria had been applied.

Application of these stratification criteria to the defined universe produced the arrangement of strata summarized in Table 3. In the 90-100% minority category there were enough students for 29 strata of 30,000 students each. Students in 90-100% minority schools constituted 58% of the defined universe. These 29 strata were established by arranging all 90-100% minority schools in a geographic-urban order and then counting off groups (strata) of approximately 30,000 students. Boundaries were placed so that schools were not divided between strata; however, districts were frequently represented in more than one stratum. The geographic-urban order followed for the first 29 strata was:

Table 3. Summary of Strata of Schools for Restandardization Sample

Strata (30,000 Pupils Each)*	Percent Minority in Individual Schools	Dominant Geography of Schools
1 - 3	90 - 100%	West Coast Metropolitan
4 - 5	90 - 100%	Southwest Metropolitan
6 - 12	90 - 100%	Midwest Metropolitan
13 - 18	90 - 100%	Northeast Metropolitan
19 - 21	90 - 100%	Mid-Atlantic Metropolitan
22 - 24	90 - 100%	Southeast Metropolitan
25 - 26	90 - 100%	Southwest Non-metropolitan
27	90 - 100%	Mid-Atlantic Non-metropolitan
28	90 - 100%	South Medium
29	90 - 100%	South Small
30 - 32	80 - 89 %	Metropolitan
33 - 34	80 - 89 %	Non-metropolitan
35 - 37	70 - 79 %	Metropolitan
38 - 39	70 - 79 %	Non-metropolitan
40 - 42	60 - 69 %	Metropolitan
43 - 44	60 - 69 %	Non-metropolitan
45 - 47	50 - 59 %	Metropolitan
48 - 50	50 - 59 %	Non-metropolitan

* There are 50 strata. Lines between groups of strata indicate shifts in geography (single line) or in percent minority (double lines).

1. Metropolitan, West to East (strata 1-24)

2. Non-metropolitan, non-Southern, West to East (strata 25-27)

4. Non-metropolitan, Southern, medium (stratum 28)

5. Non-metropolitan, Southern, Small (stratum 29)

A similar geographic-urban order was followed for the remaining 21 strata. However, as Table 3 shows, there were only five or six strata for each minority percentage type below 90%, so that geographic-urban homogeneity of equally-sized strata became more difficult to achieve. Nevertheless, each stratum was made up of schools that were rather similar to each other in the stratification criteria used: percent minority, degree of urbanization, and geographic location. According to the numbers of strata occupied by the five different percent minority categories, approximately 84% of the students in the universe were from minority groups.

4. Selection of Schools

After all schools in the universe had been distributed into the 50 strata described above, two schools were randomly selected from each stratum, using a table of random numbers. In order to verify that the 100 schools so selected on the basis of 1971 data were still in the defined universe, figures for each selected school were checked against 1972 OCR Survey forms. Two of the 100 schools were discovered to have changed characteristics sufficiently to fall outside the universe. The two schools were deleted from the sample, and substitutes were drawn at random from the same stratum.

After this adjustment, the 100 schools in the sample were tabulated by district. Forty-nine districts were represented by one school each, eleven by two schools, one by three schools, one by four schools, two by six schools, and one by ten schools. Thus 65 separate districts were included in the sample.

On March 23, 1973, after receiving USOE approval of the selected sample of schools, the following letters signed by the Assistant Commissioner for Planning, Budgeting, and Evaluation were mailed:

- Sixty-five letters to the superintendents of the districts containing the 100 selected schools. These letters gave background information and requested cooperation in the restandardization activity.

- Twenty-four letters to the chief school officers of the States in which districts had been selected for restandardization testing. Each letter enclosed a copy of the letter to superintendents, along with a list of the districts and schools selected in the State. In each State, a copy was also sent to the coordinator of the State's Committee on Evaluation and Information Systems.

- Eight letters to the commissioners of the HEW Regions in which districts had been selected for restandardization testing. Each letter enclosed a copy of the letter to superintendents, along with lists of the districts and schools selected in the Region. In each Region, a copy was also sent to the BEEO Regional senior program office.

Sample copies of these letters are provided in Appendix A of this report.

The March 23 letters to superintendents requested a phone call to the USOE Project Officer naming a contact for concluding arrangements. In most instances, replies were favorable, straightforward, and reasonably prompt. In these cases, SDC called back each contact to verify permission, to inquire about standardized tests being given this year in the selected schools, and to determine whether complementary schools were needed.

The first 30 districts contacted were asked about tests being used, in order to judge whether the California Achievement Test (CAT) battery chosen for re-standardization would have recently been given to the same students. Only six districts indicated that the CAT was being used this year in the schools selected for the sample; in most of these cases, only one grade was being tested. Other tests frequently mentioned were the Comprehensive Test of Basic Skills (9 mentions), the Metropolitan Achievement Test (6 mentions), and the Iowa Test of Basic Skills (4 mentions).

Of the 100 schools selected, 17 refused to permit testing. Reasons for failure to cooperate were varied, and included dislike of testing, dislike of ESAA, recent testing in the school selected, and disagreement with restandardization objectives. Substitutes were arranged for 16 of the 17; testing took place in 99 schools. Nine of the substitutes (in five districts) were arranged in the same districts as the originally selected schools; the other eight schools were arranged in seven other districts. In all cases, substitutes were chosen from the same stratum as the original school and were as close as possible in measure of size to the original school.

Several of the large school districts drawn into the sample occupied one or more entire strata. Substitutes in these cases could only be made within the same district (following the rule that substitutes would always be drawn from the same stratum as the originally selected school). This situation led to prolonged and only partially satisfactory negotiations with one school district; hence, the number of schools tested fell one short of the desired 100. Table 4 gives the distribution by State of districts and schools in the sample, after all substitutions had been made.

Table 4. Number of Districts and Schools in the Final Restandardization Sample

State	Number of Districts	Number of Schools
Alabama	4	4
California	5	12
District of Columbia	1	3
Florida	5	8
Georgia	2	2
Illinois	1	6
Indiana	1	1
Louisiana	1	2
Maryland	1	2
Massachusetts	1	1
Michigan	1	2
Mississippi	3	3
Missouri	2	3
New Jersey	7	8
New York	2	10
North Carolina	3	3
Ohio	3	4
Pennsylvania	2	3
South Carolina	5	6
Tennessee	1	1
Texas	10	12
Virginia	2	2
Wisconsin	1	1
TOTAL	64	99

5. Sampling Students Within Schools

Because of time constraints prior to testing, it was not feasible to get a list of students from each school in the sample so that students could be randomly selected at a central location. Procedures were needed to avoid bias in the selection process. Using one intact class as a sample of a grade composed of more than one class would have been statistically inefficient. Also, non-randomness might have entered if selection were done by persons not acquainted with sampling pitfalls.

The solution adopted contained provisions for a variety of circumstances. If, for example, the total number of students in all the school's classes at a given grade level was less than 31, then all students in that grade were selected. When the number in a grade exceeded 30, students were systematically selected from separate classes according to their last initials, as determined by an alphabetic scheme which rotated letters in alternating patterns from class to class. Procedures were also developed to randomly increase or decrease the sample for a given grade in a school in order to keep the number tested close to 30 per grade.

In summary, testing actually took place in 64 of the 691 districts identified as the appropriate population of ESAA-eligible minority-isolated schools. Ninety-nine separate school units were involved. A total of 8,999 students, approximately evenly divided from the three target grade levels, were tested. This sample represents a two-stage sampling strategy where each student in the estimated population of 1,506,751 eligible students had nearly the same chance of being in the sample.

B. TEST ADMINISTRATION

A second instrument was administered at the same time as the achievement measure to gather information related to the background of the students sampled. This instrument, the Student Background Questionnaire, asked for information related to ethnic self-identification, measures of socioeconomic level (as reflected in household possessions), language spoken in the home, and home educational experiences. A copy of this questionnaire is included as Appendix B.

Test forms were prepared on machine-scannable forms and were coded with school identification, student name, and a student identification number. The student questionnaire was also coded with student name and identification number for the purpose of matching responses on the two instruments for later analysis.

Students in 99 elementary schools were tested during May and June 1973. In order to create a standardized testing situation, the actual administration of the achievement measures was subcontracted to American College Testing Program. Professional test administrators were used to ensure greater uniformity of administration and stricter adherence to time schedule than might be possible using classroom teachers or other school personnel.

American College Testing (ACT) conducted special training sessions in which test supervisors were introduced to the materials and procedures which were to be used in the administration. Supervisors were selected who were familiar with the region and had prior testing experience. After training, each supervisor contacted individuals from the local communities of each test site to serve as proctors during testing. Supervisors and their proctors were selected to reflect as closely as possible the ethnic balance of the local school populations. This team then conducted the entire administration in a standardized setting. An example of a typical administration is given below:

Typical Administration Schedule

8:00 a.m.	Supervisors and proctors arrive at the school.
8:15 - 9:15	Supervisors review testing procedures and brief proctors on last minute details. Testing facilities are checked. Student sample is selected and students are called to test site. Introductions.
9:15 - 11:30	Student Background Questionnaire and test administration.
11:30	Approximate end of test administration. Supervisors check materials, package answer sheets, and fill out report of testing irregularities (if any).

IV. DESCRIPTIVE ANALYSIS

In order to meet the first goal of the restandardization testing (i.e., assessment of the academic needs of students in ESAA-eligible minority-isolated schools) and to describe the sample in sufficient detail to support further analysis, a descriptive analysis was initiated. This chapter first describes the sample of students actually tested. Second, characteristics of the sample, as gathered in the Student Background Questionnaire, are described. Third, the achievement levels of the sample are reported. Finally, student achievement levels and background characteristics are summarized.

A. STUDENT SAMPLE

The original sampling design called for the testing of approximately 9,000 students, 3,000 in each of the three levels of interest. Within the 99 schools actually participating, the breakdown of the 8,999 students tested is shown in Table 5.

Table 5. Achievement Test and Background Questionnaire Respondents

Grade	Achievement	Background Questionnaire	Combined
3	3,025	2,500	2,500
4	3,011	2,461	2,461
5	2,963	2,461	2,461
TOTAL STUDENTS	8,999	7,422	7,422

The smaller number of Student Background Questionnaires results primarily from the refusal of some school districts to administer that instrument, although there were several cases of individual non-respondents within schools where the questionnaire was administered. In particular, several school districts in California declined to participate in the questionnaire portion of the restandardization testing.

One large city district allowed certain of the items to be administered during the testing session, and the rest to be administered only via a questionnaire

mailed to the students' homes after the testing. These responses were then included in the data base and the school was counted as having cooperated.

Additional testing irregularities occurred in certain grade levels in three schools. The net effect of these irregularities was to disrupt the desired standardized testing situation. Since the results of these disturbances were unpredictable, the data from these grades for the three schools were removed from further analysis.

In order to ascertain whether the students who responded to the questionnaire were systematically different from those who did not, the two groups were compared on the only measure available which was common to both groups--the achievement test. Performance on each subtest was compared for the two groups. As is seen in Table 6, differences in achievement levels between the two groups were not significant. For this reason, and because it was desirable to be able to associate student background with performance, all subsequent analyses were performed on that subset of students for whom both achievement tests and background questionnaires were available.

Table 6. Comparison of Means for Total Sample and Subsample Responding to Background Questionnsire

Test	3rd grade		4th grade		5th grade	
	All students	Q respondents	All students	Q respondents	All students	Q respondents
Reading:						
Vocabulary	28.04	28.02	14.23	14.28	18.22	18.29
Comprehension	22.11	22.41	14.58	14.71	18.25	18.25
Mathematics:						
Concepts	18.37	18.48	9.39	9.44	12.42	12.41
Computation	49.34	49.82	20.40	20.61	27.06	27.21
N*	2974	2500	2927	2461	2902	2461

*Sample sizes indicated under "All students" include data from all schools in which there were no testing irregularities. Sample sizes indicated by "Q respondents" refer to those students who responded to both the achievement measure and the Student Background Questionnaire.

B. STUDENT BACKGROUND CHARACTERISTICS

The twelve items on the Student Background Questionnaire focus, for the most part, on characteristics of the student's home environment. These characteristics include the number of people the student lives with, the language spoken at home, whether or not there is someone at home who can and actually does help him/her with school work, the availability of various types of reading materials, appliances, and convenience items, and the number of hours per day the student spends watching television. In addition, there are a few questions concerned with the ethnic self-identification of the student and his/her prior educational history (i.e., the number of different schools attended and the length of attendance at the present school.)

The marginals for all items on the Student Background Questionnaire were computed for each of the three grades in the restandardization sample. Since the pattern of responses and the views expressed by the majority of the students in each grade were quite similar, the students' background characteristics for the restandardization sample are presented here in a general, descriptive manner with notable exceptions mentioned. Item-by-item breakdowns of responses are given in Appendix B for all grades combined and for each grade level. Appendix C contains item-by-item breakdowns of responses across all grade levels for each of the four modified ethnic categories defined below. Significant differences ($\alpha = .01$) in response patterns between ethnic groups were found for all items on the Student Background Questionnaire except for question #7 and part G (tape recorder) on #11. These differences are described in Table 7.

The original ethnic groups included in item #9 were Black (Negro), Oriental (Japanese, Chinese, etc.), American Indian, White, and Other (Eskimo, Hawaiian, etc.). Item #10 also asked the students if they considered themselves to be of a Spanish background (Mexican, Cuban, Puerto Rican, Latin American). Since the original marginals indicated that there were only 1.2% Oriental students and 3.4% American Indian students, these categories were merged with the "Other" category for purposes of further analysis. Also, Spanish-background students were considered

Table 7. Significant Differences Between Groups on Student Background Questionnaire Items (Sheet 1)

Item #	Differences Between Groups
1	Whites are more often new to the school. Blacks and Spanish-background students are more likely to have been at their present school since kindergarten or first grade.
2	Black, Spanish-background, and Other students more often have gone to only one school since kindergarten than have White students. White students are more likely to have gone to three schools.
3	White students are more likely to live with three to four other people at home than are Black, Spanish, and Other students. These students are more likely to live with seven or more other people than are White students.
4	Very slight difference between the ethnic groups. White and Other students tend to have done more reading at home (not related to school work) in the past two weeks.
5	White and Other students are more likely to have done school work at home during the past two weeks than are Black or Spanish-background students.
6	Spanish students are least likely to have anyone in their homes who can help them with their school work.
7	No significant differences.
8	Black students are most likely to speak only English at home, with White students slightly less likely. Other, and particularly Spanish students tend to speak another language at home besides English--Spanish for the Spanish-background students and Chinese, an American Indian language, or some other language for the Other students.
11A	Other and Spanish-background students are less likely to have a daily newspaper in their homes.
11B, C, E, H	Whites are most likely and Spanish-background students are least likely to have a dictionary, encyclopedia, magazines, and color television in their homes.

Table 7. Significant Differences Between Groups on Student Background Questionnaire Items (Sheet 2)

Item #	Differences Between Groups
11D	White students are most likely and Black students slightly less likely to have story books in their homes, while Spanish-background students are least likely.
11F	Black students are most likely to have a record player in their homes, with Other students least likely.
11G	No significant differences.
11I	White and Other students are most likely to have a typewriter in their homes. Spanish-background students are least likely to have one in their homes.
11J	White students are most likely and Black students are least likely to have a dishwasher in their homes.
11K	White students are more likely than any of the other groups to have two or more cars or trucks that run.
11L	White students are most likely and Black and Spanish-background students least likely to have an automatic clothes dryer in their homes.
11M	Other students are most likely and Spanish-background students are least likely to have a special place to study.
12	Black students are most likely to watch television more than three hours a day.

to be an important separate ethnic group. Although many students were of Spanish background (as they indicated in item #10), their responses to item #9 were scattered throughout the five ethnic groups. (The percentages for item #9 in Appendix C illustrate this point.)

In order to make the ethnic distinctions very clear, the Spanish-background students were removed from the Black, Oriental, American Indian, White, and Other categories and were merged together to form the Spanish-background ethnic group. Thus, four principal ethnic groups resulted from this process: Black, White, Spanish-background, and Other. Table 8 gives the percentage of each of these ethnic groups in the sample, for all grades combined and by grade level. These four ethnic categories were utilized in all the analyses of the achievement data and the Student Background Questionnaire data.

In the restandardization sample, 83% were minority students. The largest single group of students were Black (60%), followed by Spanish-background students (21%), White students (15%), and Other students (3%). The validity of the sampling techniques is supported by the correspondence between the estimated (84%) and actual (83%) minority representation.

Several questions dealt with various aspects of the home environment. Most frequently (31-34%), the students live in a home with five or six other people. However, White students are more likely to live with three or four other people at home than are other groups, while minority students are more likely to live with larger families of seven or more other people. Generally, only English is spoken at home (74%). However, 68% of the Spanish-background students speak Spanish at home. For students in the Other category, Chinese (8.4%), some other language (8.0%), and an American Indian language (5.6%) are sometimes spoken.

Most of the students (75%) stated that they had done reading unrelated to school work at home in the two weeks prior to testing. White (76%) and Other (71%) students are more likely to have done school work at home in the two weeks prior to testing than are Black (66%) or Spanish-background (66%) students.

Table 8. Percentages for the Modified Ethnic Categories

Ethnic Group	Grade Level			
	Combined	3	4	5
White	15%	16%	16%	14%
Black	60%	58%	60%	62%
Spanish Background	21%	22%	21%	21%
Other	3%	4%	3%	3%

Although it appears that amost all of the students can get help at home with their school work (93%), the Spanish-background students can get help least often (88%). Fewer students (53%) had actually received help during the two weeks just prior to testing.

Several valuable resource materials and conveniences are available in the homes of the students. There are differences in response patterns between the various grade levels and the ethnic groups for this item. In all grades, the four most frequent materials, in descending frequency, are: a record player (89%), story books (83%), a dictionary (78%), and magazines (70%). The last three support the contention that students do reading at home unrelated to school work.

After the first four items, students in different grades differ in the order of frequency of certain materials in their homes. Table 9 specifies the order of frequency for each of the three grades.

Students in all three grades agree on the four items that are least frequently found in their homes; the lack of these items seems to indicate the socio-economic status of the sample students' families. These four items, in descending frequency, are: two or more cars or trucks that run, a typewriter, an automatic clothes dryer, and an automatic dishwasher. These items could be classified more as luxury items than as necessities.

The minority students agree that the four most frequent items in their homes are: a record player (83-91%), story books (75-85%), a dictionary (69-80%), and magazines (65-70%). The White students indicated the same order except that story books (88%) were most frequent, followed by a record player (87%). After these items, the frequency of the next seven items differs for all the races.

Table 9. Order of Frequency of Materials in the Home by Grade Level

Order of Frequency	Grade 3	Grade 4	Grade 5
1	Record player	Record player	Record player
2	Story books	Story books	Story books
3	Dictionary	Dictionary	Dictionary
4	Magazines	Magazines	Magazines
5	Special place to study	Daily newspaper	Daily newspaper
6	Daily newspaper	Encyclopedia	Encyclopedia
7	Color TV	Tape recorder	Color TV
8	Encyclopedia	Color TV	Special place to study
9	Tape recorder	Special place to study	Tape recorder
10	Two or more cars or trucks	Two or more cars or trucks	Two or more cars or trucks
11	Typewriter	Typewriter	Typewriter
12	Automatic clothes dryer	Automatic clothes dryer	Automatic clothes dryer
13	Automatic dishwasher	Automatic dishwasher	Automatic dishwasher

Table 10 presents the order of frequency for each of the ethnic groups. Items such as a color TV and two or more cars that run are more frequent in the homes of White students than in the homes of minority students. All groups seem to have relatively easy access to a daily newspaper and an encyclopedia.

For all ethnic groups the two least frequent items are an automatic clothes dryer (27-46%) and an automatic dishwasher (10-23%). Although these are luxury items, a much larger percentage of White students respond that these items are present in their homes than do minority students.

In general, White students are more likely to have each of the items listed in question #11 in their homes, whereas Spanish-background students are the least likely.

Students frequently responded (50%) that they watch television more than three hours per day. The Black students are most likely to watch television this amount. The least frequent response was no television-watching at all (7%). The fourth- and fifth-grade students reported watching more television than the third graders.

Finally, minority students (28-31%) are more likely than White students (19%) to have gone to the same school since pre-school or kindergarten. White students are more likely (24%) to be new this year at the school. In correspondence to item #1, the results of item #2 indicate that minority students are more likely (41-46%) than White students (32%) to have gone to only one school since kindergarten.

Table 10. Order of Frequency of Materials in the Home by Ethnic Group (Combined Grades)

Order of Frequency	Black	White	Spanish	Other
1	Record player	Record player	Record player	Record player
2	Story books	Story books	Story books	Story books
3	Dictionary	Dictionary	Dictionary	Dictionary
4	Magazines	Magazines	Magazines	Magazines
5	Daily newspaper	Encyclopedia	Daily newspaper	Special place to study
6	Encyclopedia and special place to study (tie)	Daily newspaper	Encyclopedia	Daily newspaper
7	Encyclopedia and special place to study (tie)	Color TV	Tape recorder and color TV (tie)	Encyclopedia
9	Color TV	Two or more cars	Tape recorder and color TV (tie)	Tape recorder
10	Typewriter	Tape recorder	Two or more cars	Typewriter
11	Two or more cars	Typewriter	Typewriter	Two or more cars
12	Automatic clothes dryer	Automatic clothes dryer	Automatic clothes dryer	Automatic clothes dryer
13	Automatic dishwasher	Automatic dishwasher	Automatic dishwasher	Automatic dishwasher

C. STUDENT ACHIEVEMENT LEVELS

The distribution of subtests and total scores for students included in all subsequent analyses are summarized in their raw score form in Table 11.

One should take care to note that different levels of the achievement measure, containing different numbers of items, are used in grade 3 and grades 4-5. Achievement levels for third-grade students are more typically toward the upper end of the achievement levels of their fourth- and fifth-grade schoolmates. The range and skewness values for each subtest and total scores are presented in Table 12. These data may indicate the presence of a possible ceiling effect for third-grade students in the Level 2 instrument (three of four subtests are negatively skewed). The overall proportion of correct responses is only .63, suggesting that this danger is not too great.

The data also indicate considerable variability of scores within the ESAA-eligible sample, as indicated by the standard deviations and ranges reported in Tables 11 and 12. While some students are operating below chance level, others are exhibiting near-perfect performance. Score reliability as computed using Kuder-Richardson Formula 20 (KR-20), particularly for total scales, is quite adequate. Reliability values range from 0 to 1 and are an indication of the homogeneity of the items within a test and the replicability of results of measurements made at different points in time. High reliabilities (greater than .90) indicate stable measurements.

Table 11. Summary of Raw Score Test Characteristics for Restandardization Sample

SCALE	LEVEL 2					LEVEL 3								
	Grade 3					Grade 4					Grade 5			
	# Items Level 2	n	$\bar{x}$	S.D.	KR-20	# Items Level 3	n	$\bar{x}$	S.D.	KR-20	n	$\bar{x}$	S.D.	KR-20
Reading Total	85	2500	50.42	18.18	.945	82	2461	28.99	12.83	.899	2461	36.54	14.53	.916
Vocabulary	40	2500	28.02	8.70	.900	40	2461	14.28	7.09	.844	2461	18.29	8.11	.871
Comprehension	45	2500	22.41	11.03	.912	42	2461	14.71	6.81	.812	2461	18.25	7.44	.834
Math Total	102	2500	68.29	22.13	.959	73	2461	30.04	13.90	.926	2461	39.62	15.00	.935
Computation	72	2500	49.82	17.13	.954	48	2461	20.61	9.80	.903	2461	27.21	10.61	.915
Concepts	30	2500	18.48	6.55	.858	25	2461	9.44	5.13	.813	2461	12.41	5.44	.830
Battery	187	2500	118.72	37.24	.973	155	2461	59.04	24.67	.949	2461	76.17	27.10	.956

Table 12. Range and Skewness of Achievement Test Scores by Grade Level

	Grade 3		Grade 4		Grade 5	
	Range	Skewness*	Range	Skewness	Range	Skewness
Reading Total	83	0.005	78	0.723	81	0.525
Vocabulary	38	-0.502	38	0.576	39	0.359
Comprehension	45	0.199	40	0.637	42	0.503
Math Total	102	-0.598	71	0.318	72	-0.217
Computation	72	-0.692	48	0.301	48	-0.186
Math Concepts	30	-0.372	25	0.226	25	-0.279

*A distribution is considered skewed when there is a considerably larger number of extreme cases on one side of the distribution curve than on the other. When the result is a positive number, the distribution is skewed to the right (extremely high scores are farther away from the mean than are low scores); when the result is negative, the distribution is skewed to the left.

D. SUMMARY OF STUDENT ACHIEVEMENT LEVELS AND BACKGROUND CHARACTERISTICS

The data presented above describe the background characteristics and achievement levels of students sampled from ESAA-eligible minority-isolated schools. Inasmuch as the students were randomly sampled within schools across classrooms, these data should accurately reflect the characteristics and performances of students from the defined population.

The general picture that one obtains is a student population with a high percentage of minority students from lower socioeconomic levels (as indicated by the absence of major appliances). The home environments of these students include several media that could be of educational value: record players, books, dictionaries, and magazines.

The achievement levels of these students, when compared to the national norms for each measure*, indicate depressed levels of performance. Table 13 shows the percentile rank and grade-equivalent (GE) level associated with the median performance for each grade on each subtest. Percentile ranks typically hover around the twentieth percentile level. The students tend to be somewhat weaker in reading and math concepts than in mathematics computation performance. Additionally, these students tend to fall further behind grade level as they advance through school. The latter result indicates that these students reap less than a full year of learning during a year of schooling.

The need of students in minority-isolated schools is clearly established by these data. Students eligible for compensatory funds under the definition of ESAA actually are achieving at depressed levels. In addition to establishing need, these data can be used as a baseline for comparisons with later evaluation data.

*Because the mathematics subtests used differ from those originally normed, special norms were requested and obtained from the test publisher based on the same data and process as the original norms.

Table 13. Achievement Levels of Restandardization Sample Relative to National Norms

Reading Vocabulary				Reading Comprehension			Total Reading			
Grade	Median	Percentile	GE	Median	Percentile	GE	Median	Percentile	GE	GED*
3	29.687	25	2.7	21.020	21	2.6	50.033	23	2.7	-1.2
4	13.360	22	3.0	13.681	18	3.1	26.849	18	3.0	-1.9
5	17.485	21	3.9	17.351	18	3.7	34.693	19	3.9	-2.0

Math Concepts				Math Computations			Total Math			
Grade	Median	Percentile	GE	Median	Percentile	GE	Median	Percentile	GE	GED
3	19.210	18	2.4	53.159	31	3.2	7.430	25	3.0	- .9
4	8.922	14	2.8	19.455	17	3.5	28.600	15	3.4	-1.5
5	12.912	20	3.9	27.911	23	4.5	40.550	20	4.3	-1.6

*Grade Equivalent Difference (GED) is defined as the difference between actual grade equivalents of mean scores and the expected level of performance.

An analysis of the achievement data by ethnic self-identification indicates that performance levels are not only generally depressed but are differentially depressed for different subgroups of students. Table 14 shows the achievement means, standard deviations, and grade equivalents, by ethnic category and grade level, for each subtest. An invariable ordering of performance level is present, with White students always scoring highest, Spanish-background next, and Black students scoring lowest. For the Level 2 measure (Grade 3) Blacks are more homogeneous (less variability in test scores) in reading, while Whites are more homogeneous in mathematics. For Level 3 (Grades 4 and 5) a similar pattern exists for reading performance, but mathematics score variability changes with grade level. At the fourth grade, Black students are the most homogeneous group, while at the fifth grade, White students are the most homogeneous in mathematics achievement.

Inspection of the grade equivalents* associated with each of these achievement levels indicates some interesting patterns. While all students in ESAA-eligible minority-isolated schools tend to be achieving below grade level and falling more and more behind as they progress through school, there are differential patterns of this phenomenon between ethnic groups. Grade equivalents are expressed in years and months; thus the expected performance at the end of the third grade should be 3.9, at the end of the fourth grade should be 4.9, and at the end of the fifth grade should be 5.9. While all students are falling behind more each year, White students are doing so least rapidly, Black students most rapidly, and Spanish-background students somewhere in between. This indicates that differences between minority and White students are tending to increase as the students move from one grade to the next.

*Because these grade equivalents are based on mean performance level, they are somewhat different from those based on the medians in Table 13. The medians are the most appropriate indicator of typical performance, but performance patterns are the same for both indicators.

Table 14. Achievement Test Performance by Ethnic Self-Identification (Sheet 1)

Data for Grade 3

Scale	Group	Mean	Standard Deviation	Grade Equivalent	N
Vocabulary	Other	28.670	8.919	2.6	(91)
	White	31.608	8.346	2.8	(385)
	Spanish Background	28.429	8.656	2.5	(532)
	Black	26.857	8.497	2.4	(1447)
	Total	28.0098	8.5192	2.5	(2455)
Comprehension	Other	22.835	11.138	2.7	(91)
	White	27,610	11.320	3.1	(385)
	Spanish Background	22.008	11.108	2.7	(532)
	Black	21.106	10.492	2.6	(1447)
	Total	22.3853	10.7781	2.7	(2455)
Reading Total	Other	51.505	18.495	2.7	(91)
	White	59.218	18.396	3.0	(385)
	Spanish Background	50.436	18.155	2.7	(532)
	Black	47.963	17.323	2.6	(1447)
	Total	50.3951	17.7102	2.7	(2455)
Math Concepts	Other	18.549	6.509	2.3	(91)
	White	21.860	5.971	2.9	(385)
	Spanish Background	18.953	6.471	2.4	(532)
	Black	17.356	6.380	2.1	(1447)
	Total	18.4525	6.3385	2.3	(2455)
Math Computation	Other	48.319	17.089	2.9	(91)
	White	57.262	15.173	3.4	(385)
	Spanish Background	50.711	16.950	3.1	(532)
	Black	47.641	17.085	2.9	(1447)
	Total	49.8399	16.7602	3.0	(2455)
Math Total	Other	66.868	21.771	2.8	(91)
	White	79.122	19.553	3.4	(385)
	Spanish Background	69.664	21.927	2.9	(532)
	Black	64.997	21.910	2.7	(1447)
	Total	68.2924	21.5435	2.8	(2455)

Table 14. Achievement Test Performance by Ethnic Self-Identification (Sheet 2)

Data for Grade 4

Scale	Group	Mean	Standard Deviation	Grade Equivalent	N
Vocabulary	Other	14.692	7.337	3.3	(78)
	White	17.868	7.877	4.0	(379)
	Spanish Background	14.240	6.739	3.2	(496)
	Black	13.202	6.549	3.0	(1465)
	Total	14.1944	6.8335	3.2	(2418)
Comprehension	Other	15.295	7.419	3.3	(78)
	White	18.150	7.795	3.9	(379)
	Spanish Background	14.633	6.740	3.2	(496)
	Black	13.762	6.135	3.0	(1465)
	Total	14.6778	6.5846	3.2	(2418)
Reading Total	Other	29.987	13.817	3.3	(78)
	White	36.018	14.900	4.0	(379)
	Spanish Background	28.873	12.274	3.2	(496)
	Black	26.964	11.476	3.0	(1465)
	Total	28.8722	12.3056	3.2	(2418)
Math Concepts	Other	9.038	5.810	2.8	(78)
	White	12.309	5.322	3.7	(379
	Spanish Background	9.643	5.051	2.9	(496)
	Black	8.618	4.745	2.7	(1465)
	Total	9.4202	4.9361	2.9	(2418)
Math Computation	Other	19.949	10.502	3.6	(78)
	White	25.156	10.453	4.2	(379)
	Spanish Background	21.802	10.208	3.8	(496)
	Black	19.091	9.021	3.5	(1465)
	Total	20.6257	9.5518	3.7	(2418)
Math Total	Other	28.987	15.197	3.4	(78)
	White	37.464	14.898	4.0	(379
	Spanish Background	31.446	14.269	3.5	(496
	Black	27.709	12.643	3.3	(1465)
	Total	30.0459	13.4380	3.5	(2418)

Table 14. Achievement Test Performance by Ethnic Self-Identification (Sheet 3)

Data for Grade 5

Scale	Group	Mean	Standard Deviation	Grade Equivalent	N
Vocabulary	Other	20.420	8.874	4.6	(81)
	White	23.474	8.364	5.1	(346)
	Spanish Background	18.070	8.154	4.0	(502)
	Black	17.045	7.475	3.8	(1499)
	Total	18.2858	7.7947	4.1	(2428)
Comprehension	Other	20.444	8.191	4.5	(81)
	White	23.055	8.228	5.0	(346)
	Spanish Background	18.683	7.439	4.0	(502)
	Black	16.874	6.667	3.6	(1499)
	Total	18.2479	7.1200	3.9	(2428)
Reading Total	Other	40.864	16.049	4.6	(81)
	White	46.529	15.764	5.1	(346)
	Spanish Background	36.753	14.668	4.1	(502)
	Black	33.919	12.962	3.8	(1499)
	Total	36.5338	13.8538	4.1	(2428)
Math Concepts	Other	12.975	5.604	3.9	(81)
	White	15.344	5.121	4.9	(346)
	Spanish Background	12.588	5.530	3.8	(502)
	Black	11.661	5.246	3.5	(1499)
	Total	12.4213	5.2972	3.7	(2428)
Math Computation	Other	26.704	10.537	4.4	(81)
	White	31.173	9.816	4.9	(346)
	Spanish Background	28.534	11.255	4.6	(502)
	Black	25.895	10.328	4.3	(1499)
	Total	27.2199	10.4571	4.4	(2428)
Math Total	Other	39.679	15.168	4.2	(81)
	White	46.517	13.940	4.9	(346)
	Spanish Background	41.122	15.796	4.3	(502)
	Black	37.556	14.465	4.1	(1499)
	Total	39.6413	14.6919	4.2	(2428)

In addition to looking at the relationship between ethnic group and achievement scores, attention was also given to the possible impact of socioeconomic status (SES) on achievement levels, and its relation to ethnicity. Although no clear-cut measure of SES was available, a rough indicator of socioeconomic status was available in the form of the item that asked whether or not students had certain specified resources in their homes. These resources clustered into two groups--those that were more educationally-oriented (such as books) and those that were relatively expensive luxury appliances (such as a dishwasher). Thus two SES scales were derived, each consisting of five items, and a score was computed for each student representing the number of such items (ranging from 0 to 5) in his home. Specifically, the first SES scale was comprised of the following five items: daily newspaper, dictionary, encyclopedia or other reference books, story books, and magazines. The items on the second scale included: tape recorder or cassette player, typewriter, automatic dishwasher, two or more cars or trucks that run, and an automatic clothes dryer. Overall, the average number of items in the first group that the students had in their homes was 3.55, and in the second group it was 1.90. Not surprisingly, both of these SES scales were significantly related to ethnic group; in both cases, White students were likely to have the most of these items and students of Spanish background the least. Tables 15 and 16 show, for each scale, the percentage distribution and mean score on the scale for each ethnic group and for the total sample.

Next, the possible impact of SES on achievement level was explored. Each of the two SES measures was related to the student's total reading and total math scores for each of the three grade levels under study. At all grade levels, and for both reading and math, each of the SES scales showed a significant positive relationship with achievement score. In all cases, the first SES measure, comprised of educationally-oriented items, was more strongly related to achievement than was the second SES scale. Also, for both SES measures, there was a stronger relationship with reading than with math scores.

Table 15. Ethnic Group and First SES Measure

Scores on First SES Measure (number of educationally-oriented items in home)	Ethnic Group				
	Black	Spanish Background	White	Other	Sample
0	2.4%	4.7%	1.7%	4.4%	2.8%
1	6.0	10.2	4.3	10.4	6.8
2	11.4	17.7	9.2	10.4	12.4
3	21.4	21.0	16.1	16.4	20.4
4	27.9	21.8	26.7	25.6	26.3
5	31.0	24.5	42.0	32.8	31.4
	100.0%	100.0%	100.0%	100.0%	100.0%
Mean Score	3.59	3.19	3.88	3.47	3.55

Table 16. Ethnic Group and Second SES Measure

Scores on Second SES Measure (number of appliances in home)	Ethnic Group				
	Black	Spanish Background	White	Other	Sample
0	17.8%	20.3%	12.6%	16.8%	17.5%
1	26.7	28.0	17.8	23.2	25.5
2	25.5	22.7	24.0	26.0	24.7
3	18.4	15.8	21.5	20.0	18.4
4	8.7	8.6	14.1	8.0	9.5
5	3.0	4.7	9.9	6.0	4.5
	100.0%	100.0%	100.0%	100.0%	100.0%
Mean Score	1.82	1.79	2.36	1.97	1.90

Finally, the impact of SES on the previously discussed relationship between ethnic group and achievement level was explored, to determine whether this relation was merely, or largely, a function of different SES levels in the different ethnic groups. To explore this, an analysis of covariance of ethnic group on achievement level was done, using the two SES measures as covariates. The results of this analysis showed that, while SES had significant impact on achievement score, the reading and math scores of the different ethnic groups remained significantly different even after adjustment for the effect of SES level. These same results were found in all three grades, with patterns of differences being identical to those previously reported for unadjusted scores. A typical result is shown in Table 17, indicating that although the two SES measures are significantly related to total reading scores for fourth graders (as indicated by a significant F value for the regression slope), there still exists a significant difference between ethnic groups on adjusted scores.

Table 17. Analysis of Covariance for Fourth-Grade Total Reading Scores

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	F-Value
Ethnic Group (Adjusted)	3	19441.0	6480.3	45.49*
Slope	2	22343.3	11171.7	78.42*
Error	2412	343626.6	142.5	

*P<.01

V. DEVELOPMENT OF UNBIASED ESAA ACHIEVEMENT MEASURES

A. RATIONALE AND METHODOLOGY

While bias is an emotionally charged term, it has a straightforward technical meaning. A test is biased if it measures different things for identifiable subgroups in the population. Often these subgroups are defined along cultural and ethnic lines. The fact that different groups attain different average scores does not of itself indicate bias. The bias occurs when the scores are used for comparing the groups in an inappropriate way. If, for example, a test purporting to measure reading comprehension is administered to two groups, and for one group, perhaps because of their cultural background, the test is more of a vocabulary measure, then the bias occurs when one tries to compare members of the two groups on the reading comprehension dimension. The test as a measure of reading comprehension is biased against the group for which it is primarily a vocabulary measure. However, if the test is used for comparing individuals within the second group on the vocabulary content, then the test is not biased, since it measures the same content for all members of this group and is used appropriately.

It is not surprising that many investigators have found evidence of bias against cultural and ethnic minorities in popular aptitude and achievement measures. These measures have been typically developed by and for "middle America" and reflect the content that is thought appropriate to this group. If these measures favor any group, they favor White middle-class students. To the extent that definable subgroups share life experience with that group, the test is appropriate. To the extent that the content reflects factors that are unique to a particular culture or have culturally specific meanings, the test is biased in favor of members of that culture and against members of other cultures. Williams (1974), for example, has developed reading passages that bias reading comprehension tests in favor of Black students, reversing the more typical bias in favor of White middle-class students.

A number of methods for investigating test bias have been suggested (Cardell and Coffman, 1964; Cleary and Hilton, 1968; Sharon and Angoff, 1973; Green and Draper, 1972). In the absence of an external validity criterion*, all such methods are concerned with the detection of items that show differential characteristics in defined subgroups. Two of these approaches have been used extensively. One approach investigates the items that contribute to the group-by-item interaction within an analysis-of-variance framework. The other focuses on methods for maximizing certain psychometric properties of subsets of items through item analysis techniques, and then comparing the resultant subsets. Items that are good discriminators in one subgroup but not in others are identified as biased against those other groups.

While considerable research effort has been expended in developing these techniques, the research reported in the literature is directed more toward the statistical methodology than toward the applied problem of identifying bias in a measure and taking some corrective action. For purposes of the ESAA evaluation, it was deemed necessary not only to identify the possibly biased items, but to remove them from the measures, thereby deriving appropriate and maximally sensitive measures of achievement. It should be noted that by beginning with measures designed for White middle-class students and then removing items that demonstrate possible bias against the ESAA subgroup, one defines a measure representing the educational content and experience common to both groups.

The following sections describe the steps undertaken to identify possibly biased items. Briefly, two phases were used in the bias analysis. The first phase, a statistical analysis, identified items with statistical characteristics indicating that the items might be biased. However, since such characteristics could have resulted from random sampling fluctuations, a second phase investigated the item content more intensely. This second phase, the content analysis,

*A common method of validating a test is to compare performance on the test with some independent criterion external to the test. If the test is correlated with the other criterion then the test is said to be a valid measure.

focused on determining whether each item had a content or form that might bias it against students in minority-isolated schools. The basis for content analysis decisions was the consensus of minority-group experts on measurement theory, testing, subject matter content, compensatory education, and problems of the disadvantaged student.

It must be noted that these procedures do not guarantee the identification of truly biased items, nor do they insure that any item identified is truly biased. Rather, these procedures identify items for which the probability that the item is biased is significantly greater than zero. That is, for the items identified, the preponderance of evidence available suggests a bias. It is therefore necessary to remove those items from the scoring when an unbiased measure is desired. Their removal lessens the chance of including biased items. The procedures do not rule out the possibility that biased items are still included, and they do not in any sense "prove" that the items removed are truly biased.

B. STATISTICAL ANALYSIS

The basic data for the statistical analysis came from item-characteristic indices derived from the test publisher's national standardization sample and from the special administration to the sample of students enrolled in ESAA-eligible minority-isolated schools. (See CTB's Bulletin of Technical Data for descriptions of samples and sampling techniques for publishers' standardization.) The logic of these analyses followed the logic used in other item bias investigations in the absence of an external validity criterion. Here, however, less concern was given to determining statistical signifiance as a criterion for item bias than was placed on identifying "suspicious" items for further study in the content analysis phase. For this reason the statistical procedures were modified somewhat, in order to give a better picture of each item in relation to the other items regardless of level of significance.

In the analysis-of-variance framework for investigating bias, researchers have analyzed the data in a two-dimensional model that considers items as one dimension and group membership (e.g., ethnic group, SES level, etc.) as the

other dimension. Within this framework the investigator attempts to determine which items contribute to the item-by-group interaction; that is, which items demonstrate a difficulty level for a certain group that cannot be reasonably accounted for by the overall level of the item, the level of the group, or the general difficulty level of the test. The first procedure used here provides information directly indicating which items contributed to the item-by-group interaction, without performing the statistical tests. Each of the four subtests was first ordered by the difficulty values derived from the publisher's standardization sample. An item-by-difficulty plot was prepared. Data from the ESAA-eligible sample were then entered on the same plot. If no interaction was present, the resulting curves would be similar in shape but might have differing heights depending on the overall achievement levels. Items contributing to the item-by-group interaction appeared as disturbances in the uniformity of the curves. Similar graphs were prepared for several ethnic subgroups within the ESAA sample.

As an example, a hypothetical case is shown in Figure 1 below where items 2 and 5 show different characteristics in the two student groups (A and B). Because of such marked differences, these items are highly suspicious. The difference in the overall height of the curves indicates that the test is generally more difficult for Group B. It is of interest to note that in this example item 2 is biased against members of group B, while item 5 is biased in favor of group B.

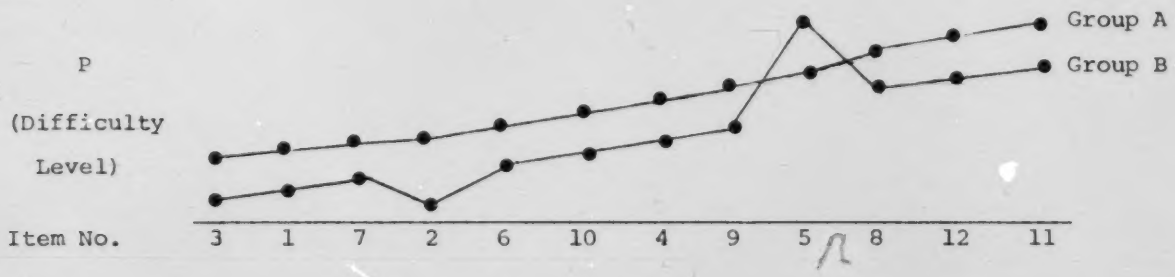

Figure 1. Example of Graphical Detection of Item-by-Group Interaction

The second approach to the statistical analysis of item bias used information normally used for test construction. When designing an achievement measure, the test constructor wants to achieve two general objectives. First, he wishes to measure one particular content; e.g., a reading vocabulary test should measure reading vocabulary knowledge and not mathematical computation skills. Second, the measure should discriminate levels of knowledge of the examinees in relation to the content. One method of analyzing a set of items under consideration for inclusion in a test is to examine the point-biserial correlation coefficients between the score on each item and the total score on the set of items. An item will exhibit a lower correlation if it measures a content unrelated to the content measured by most of the other items present. Thus, this technique can be used to look for biased items. It should be noted that low point-biserial correlations can also result from other psychometric properties of the item, but that in published tests one can assume that items with low correlations for other reasons have already been removed from the item pool.

The second procedure, then, was based on the discriminability quality of each item in the subtest. A low point-biserial correlation coefficient for an item did not contribute to total scores for the subtest and indicated that the item might be measuring a different content from that measured by other items in the subtest. By comparing the point-biserial values for the defined groups, it was possible to identify items that did not contribute to total scores for certain subgroups. Again the ESAA-eligible minority-isolated sample as a whole was compared to the standardization sample, and comparisons were made among subgroups within the ESAA-eligible sample.

Using the above procedures, the restandardization data were analyzed. In order to identify "suspicious" items for more intensive analysis, statistical significance criteria were abandoned in favor of a more subjective review of the statistical results. Three professional staff members with

considerable experience and training in statistical analysis and psychometric theory jointly reviewed these results and identified the items which in their collective opinion indicated aberrant properties.

Two kinds of patterns were typically noted. Some items had different characteristics (item difficulty or discriminability) for the ESAA-eligible minority-isolated sample in general but not necessarily among groups within that sample. For example, in Table 18, the point-biserial correlation coefficients for Item 1 in the Level 3 Vocabulary subtest for the three main groups within the ESAA sample are lower than the corresponding coefficient for the CTB standardization sample, but there are no large differences between groups within the restandardization sample. (Grade 5 is reported here, but a similar pattern exists for Grade 4.)

Other items showed considerable variability among subgroups within the restandardization sample. The second row of Table 18 illustrates this case, using Item 6 of the Level 3 Comprehension subtest (again using fifth-grade data). Here the item is a good discriminator for Whites within the restandardization sample, but the discriminability falls off for the other two major subgroups.

Table 18. Examples of Aberrant Item Characteristics

Case (Example Items)	ESAA-Eligible Sample			CTB Sample
	Black	Spanish	White	
1. Item 1, Level 3 Vocabulary	.29	.31	.27	.50
2. Item 6, Level 3 Comprehension	.35	.34	.56	.56

In all, 56 items were identified as suspicious, eight from Level 2 and 48 from Level 3. Table 19 indicates the number of suspicious items from each Level and subtest. Clearly the areas of reading skills tend to be the most heavily laden with suspicious items. The mathematics items identified at Level 3 may be the result of the extreme difficulty values of many of these items for students within the restandardization samples.

Table 19. Summary of Number of Items Identified as Suspicious, by Level and Subtest

Subtest	Level 2		Level 3	
	Total Number of Items	Possible Bias	Total Number of Items	Possible Bias
Reading				
Vocabulary	40	5	40	10
Comprehension	45	3	42	19
Mathematics				
Concepts	30	0	25	7
Computation	72	0	48	12

The items identified as statistically suspicious and submitted to content review are listed in Table 20, along with the reasons for such identification. The categories indicate the kinds of data that were considered the primary reasons for flagging a particular item. Extreme difficulty, for example, accounts for a certain subset of the items. Discriminability in both of the senses described above and contribution to the item-by-group interaction are the other categories noted. It is instructive to note that both the first and third columns are related to difficulty. In the first column, items are either too hard or too easy for a particular group or groups.

Table 20. Items Identified as Potentially Biased and Reasons for Identification (Sheet 1)

Level 2, Grade 3

Subscale	Item #	Difficulty	Discrim-inability	Interaction
Vocabulary	1	X		
	2	X		
	38		X	
	39		X	
	40		X	
Comprehension	2	X		
	3			X
	7		X	

Level 3, Grades 4 & 5

Subscale	Item #	Difficulty	Discrim-inability	Interaction
Vocabulary	1		X	
	12			X
	16			X
	17			
	23			
	29		X	
	32		X	
	38	X		
	39	X		
	40	X		

Table 20. Items Identified as Potentially Biased and Reasons for Identification (Sheet 2)

Level 3, Grades 4 & 5 (continued)

Subscale	Item #	Difficulty	Discrim-inability	Interaction
Comprehension	6		X	
	11		X	
	12		X	
	13			X
	15			X
	19		X	
	20			X
	21	X		
	22	X		
	23		X	
	28		X	
	29			X
	30		X	
	31	X		
	33		X	
	34		X	
	40		X	
	41	X		
	42	X		
Math Concepts	11		X	
	19	X		
	21	X		
	22	X		
	23	X		
	24	X		
	25	X		

Table 20. Items Identified as Potentially Biased and Reasons for Identification (Sheet 3)

Level 3, Grades 4 & 5 (continued)

Subscale	Item #	Difficulty	Discrim-inability	Interaction
Math Computation	2			X
	20	X		
	39	X		
	40		X	
	41		X	
	42		X	
	43			X
	44	X		
	45	X		
	46	X		
	47	X		
	48	X		

C. CONTENT ANALYSIS

The items identified by the above procedures were suspected of bias because of their differing statistical properties in different subgroups. However, the statistical analysis could not indicate exactly what property of an item made it biased. In fact, the statistical analysis could have identified some items that were not biased but that had aberrant characteristics because of chance factors. Since the goal for the ESAA Evaluation was to develop measures that would not contain bias against any of the subgroups involved in the study, it was considered necessary to identify the source of potential bias in an item before removing it.

The review of suspicious items was conducted to determine whether each item had a content or form that could bias it against one or more of the ESAA subgroups. If, for example, a reading comprehension item contained item alternatives requiring knowledge of a particular culture, this item would be considered potentially biased against students from other cultures. Such an item may not measure reading comprehension for members of other cultures, and hence could be biased according to our earlier definition.

In order to represent the subgroups involved in the ESAA sample and to include several perspectives on the issue of test bias, the review panel was designed to reflect both the ethnic/cultural structure of the ESAA sample and various points of view on measurement issues. Eleven panelists were selected, representing different parts of the country, ethnic subgroups, and substantive points of view. The panel included three Southern Blacks, two metropolitan Blacks, one Northeastern Puerto Rican, three Southwestern Mexican Americans, one American Indian, and one Asian American. Their specific backgrounds were diverse; several were experienced teachers, two were item construction and test development specialists, and one each was a principal, a superintendent, and a community leader.

The 11 members of the review panel convened to examine independently the content of the potentially biased test items and to rate them as biased or not biased. To accomplish this goal, the following procedures were used:

1. Explanation of Objectives and Procedures

A description of the restandardization study and its objective of establishing scales that would more appropriately reflect ESAA student achievement gains was presented. It was emphasized that the statistical analysis of the responses to the CAT items had identified a number of items that had suspicious statistical characteristics for certain groups. An explanation and examples were provided to the panelists to illustrate the important distinction between a "difficult" item and a "biased" item. A "biased" item was defined as an item having an unusual difficulty level or correlation with total scores for a certain minority group, or groups, because of cultural or socioeconomic considerations.

2. Initial Rating Procedure

Of the 56 items examined by the review panel, 48 were taken from Level 3 of the CAT (administered to fourth- and fifth-grade students) and eight were taken from Level 2 (administered to third-grade students). Each group of items was preceded by instructions from the administrator's manual and the examples that were in the test booklet. The participants were instructed to rate each item as "unbiased," "slightly biased," or "more than slightly biased." This breakdown had the advantage of eliciting slight but important ratings of bias for items that the respondent might otherwise categorize as unbiased. If an item was rated as "slightly biased" or "more than slightly biased," the rater was asked to write a specific reason why the item was seen as biased. All ratings and comments were done independently and in writing, in order to ensure complete candidness and in order to remove possible influences of stronger personality types or status. The participants were encouraged to present their objections at a level of specificity that all other participants could read and understand. They were also asked to refrain from discussing all items with the other participants until the end of the day. After all reviewers had rated the items, the results were tabulated by the research staff.

3. Initial Scoring of Items

The purpose of the content analysis was to arrive at a consensus of the reviewers on whether each item was biased, and if so, why. The tabulation, therefore, looked for consensus both across reviewers and among reviewers within ethnic subgroups. If there was consensus that a particular item was either "not biased"

or "biased," that item was removed from further review. Items for which there was disagreement were used in the second round of review.

The two biased categories were treated the same. However, the various categories of respondents were weighted, since the number of respondents representing a particular minority was roughly proportionate to the ESAA sample. The following criteria were adopted for determining the status of each item:

- There would be consensus that an item was biased if at least 50% of all the respondents rated it as biased.
- There would be consensus that an item was not biased if at least 50% of the respondents in each minority category rated it as not biased.
- There would be no consensus on an item if at least 50% of the respondents in any category rated it as biased but less than 50% of all respondents rated it as biased. This item would be presented to the participants for a second rating, with its compiled list of objections.

4. Second Rating of the Items

After the tallies of responses from the first round were complete, the resultant "no consensus" items and their respective lists of objections were submitted to the panel members for review. The participants were asked to read carefully the list of objections to each item and then to rate the item a second time. The order of presentation of the items was randomized, so that the participants could remain in a group situation without influencing each other through expressions or remarks.

If the arguments attending an item were valid, then reviewers who had previously rated the item as "not biased" were expected to shift their ratings toward a consensus position. The criterion for consensus in the first rating session was intentionally conservative in order to provide participants a second opportunity and additional information with which to rate marginal items. In the event that

no shift occurred, consensus within the subgroups or groups that had originally identified the item as possibly biased was used as the criterion for retention or deletion, since some arguments against an item might be so culturally or ethnically specific as not to have relevance for members of other subgroups.

D. RESULTS

At the end of the second round, a complete list of the items identified as possibly biased and the reasons for such identification was compiled. Of the original 56 "suspicious" items, 16 items were agreed to be biased--3 from Level 2 and 13 from Level 3. All of these items were reading items. None of the math items were considered to be biased. The items identified as biased, with the reasons given, are indicated in Appendix D.

The reasons listed for items being possibly biased were diverse and perceptive. In general, lack of experience in many different areas was the major factor. When a child's experience is confined to an inner-city ghetto, to a particular region of the country, or to a city or rural area exclusively, he learns very little outside his own community or environment. His unfamiliarity with certain objects, concepts, or words described or used in the test prevent him from being able to answer the items correctly. His knowledge could be more accurately measured by using topics and words with which he is familiar.

In some of the languages represented by the various ethnic subgroups, different meanings, connotations, and implications were introduced when an English word used as an item response alternative was translated into their own language. Also, improper associations resulted when there was not sufficient knowledge of the double meanings of many English words.

From the results of the content analysis, the final ESAA scales were then determined by including only those items which were not identified as possibly biased by either statistical or content review procedures. The psychometric properties of the resultant scales are summarized in Tables 21 and 22. Table 21 presents

Table 21. Summary of Test Characteristics

Scale	Level 2		Level 3			
	Grade 3		Grade 4		Grade 5	
	# Items	KR-20*	# Items	KR-20	# Items	KR-20
Original Scales						
Reading Total	85	.94	82	.90	82	.92
Vocabulary	40	.90	40	.84	40	.87
Comprehension	45	.91	42	.81	42	.83
Debiased Scales						
Reading Total	82	.94 (.94)	69	.90 (.91)	69	.91 (.92)
Vocabulary	37	.90 (.90)	32	.84 (.87)	32	.86 (.89)
Comprehension	NA		37	.81 (.83)	37	.83 (.85)

*Values in parentheses represent stepped-up reliabilities (Spearman-Brown).

Table 22. Test Score Means for Full and Derived Scales for Two Reference Groups

	ESAA	CTB
Grade 3		
Vocabulary	28.02	30.93
Debiased Vocabulary	26.88	29.46
Difference	1.14	1.47
Grade 4		
Vocabulary	14.28	19.70
Debiased Vocabulary	12.39	16.73
Difference	1.89	2.97
Comprehension	14.71	20.27
Debiased Comprehension	13.40	18.52
Difference	1.31	1.75
Grade 5		
Vocabulary	18.29	24.69
Debiased Vocabulary	15.94	20.71
Difference	2.35	3.98
Comprehension	18.25	24.43
Debiased Comprehension	16.70	22.20
Difference	1.55	2.23

the reliability of the full and derived measures for each of the scales affected by removing items identified as biased. As the values in parentheses indicate, the reliabilities of the derived scales, when adjusted for test length by the Spearman-Brown formula*, are just as high as the original scale. Of greater importance are the means for the ESAA and CTB samples, reported for the affected subscales in Table 22. Here one notes that the removal of the items identified as possibly biased has a significantly smaller effect on total score for ESAA-eligible students than for the publisher's standardization sample. This indicates that the items removed do in fact have a greater contribution to total scores for the publisher's sample than for the ESAA-eligible sample, and that these represent little more than measurement noise for the ESAA-eligible sample.

*The Spearman-Brown prophecy formula, which can be used to estimate the effect of an increase in test length on reliability, assumes that the items added to the test are similar to the initial items in difficulty, intercorrelations, and content. Since reliability is, in part, a function of test length, these estimates are useful for comparing the original and derived scales.

VI. PREPARATION OF NORMS

Individual achievement scores describe the number of items that the examinee answered correctly. Unless the scores have an inherent underlying scale, such as scores on a true Guttman scale, it is difficult to use these raw scores in a meaningful way. Appropriate guides are needed for the interpretation of these scores. The present section addresses the problem of providing a framework for interpreting the raw scores obtained from ESAA-eligible students.

The first step in this process is the definition of an appropriate scale structure. In order to make present results maximally compatible with the normative interpretation of scores provided by the publisher's national norms tables, it was necessary to define similar kinds of norms. Percentile ranks of individual raw scores were decided upon as an appropriate interpretive scale with maximum compatibility with the national norms. Percentile rank conversions of a single raw score to the alternative comparison distribution allow quick assessment of performance relative to these two groups without further transformations.

Examination of the raw score distributions indicated that all distributions were fairly regular in shape. They were typified by a degree of skewness. Because of the wide variability and skewness, it was decided that the smoothed raw score distributions (instead of normalized scores) would be used for determining the conversions.

Separate conversion tables were prepared for individual student scores and school mean scores. Both were derived in the same manner. The cumulative score distribution was constructed from the data (individual student scores or school means). These distributions were then smoothed to minimize the effect of local irregularities. Because of the extreme regularity of the data, a rolling weighted average procedure developed by Cureton and Tukey (1957) was employed. Working from the smoothed curve, new percentile rank values were read from the curve at the mid-point of each score interval.

Appendix E contains the tables for converting raw scores to percentile ranks for each of the subtests and totals at each of the three grade levels, for both individuals and schools. For any raw score value, these tables give the percentage of cases falling below that score. In terms of their use in the ESAA Evaluation, they indicate the approximate percentage of students (or schools) from the ESAA-eligible population who have obtained a lower score on that subset of items seen as appropriate for this population, when measured at the end of their respective grade levels.

VII. SUMMARY

This document has described a series of activities designed to meet several pre-evaluation goals of the national ESAA Evaluation study. These goals included the selection of an appropriate achievement measure, a pretest of this achievement measure to assess the needs of students in schools that would be eligible to receive funds under the Act, an important research effort directed toward possible ethnic and/or cultural bias in the achievement measures, and the establishment of achievement test norms to aid in the interpretation of student and school performance relative to the appropriate subpopulation.

Toward the achievement of these goals a substantial test review and selection activity was undertaken. This activity resulted in the selection of the specific subtests of the California Achievement Test battery that were seen as best measuring the outcomes stated as objectives in the Emergency School Aid Act. These measures were selected on the basis of several criteria including appropriate content, good pychometric quality, administrative ease, and clearly defined national norms.

A nationally representative sample of students in ESAA-eligible minority-isolated schools was selected to be tested. A standardized administration of the achievement measure, as well as a questionnaire describing students' backgrounds, yielded the data that were analyzed and reported upon in the present document. These data were used in achieving the remaining goals.

Descriptive analysis yielded important baseline data for the evaluation study and firmly established the basic educational needs of students in eligible schools. This highly concentrated minority subpopulation demonstrates achievement levels significantly below those expected for their grade level. While both reading and mathematics achievement are depressed, mathematics performance is slightly better. Results indicate that minority students in minority-isolated schools demonstrate

lower performance than their white schoolmates, even though the latter are themselves significantly below their expected performance levels.

Additional descriptive analyses provided information on the characteristics of the sample. These data, particularly ethnic self-identification and socio-economic indices, are seen as crucial for establishing the comparability of future study groups to this reference group. Such comparability is the essence of the validity of such references.

Investigation of potential ethnic or cultural bias in item content and form through statistical analysis and content review yielded a small subset of items for which there was some evidence indicating possible bias. It is suggested that for certain uses of the achievement measures, scores should be calculated without these items. Such uses would include any situation where one compares groups of students (schools, programs, etc.) that are composed of significantly different proportions of students from different ethnic groups or different SES levels.

A final product of this research is the establishment of subpopulation norms based on end-of-year performance of a representative sample of students enrolled in ESAA-eligible minority-isolated schools. Such norms may be of significant use to the local evaluator in assessing the performance of an individual student or school relative to this reference group. Reference norms are provided both for complete subtests and for subtests excluding those items identified as possibly biased, and are structured for use with either individual student scores or school mean scores.

The research reported provides important baseline and supportive data for assessing the adequacy of the measures to be used in the national ESAA Evaluation study. As a reult of this research, a clear academic need has been established for students in the defined population and appropriate references have been derived for evaluation purposes.

REFERENCES

Bulletin of Technical Data, Number 1, California Achievement Tests, CTB/McGraw Hill, Monterey, 1970.

Cardell, C. and Coffman, W.E. A method for comparing the performance of different groups on the items in a test. Research Bulletin 64-61, Educational Testing Service, Princeton, 1964.

Cleary, T.A. and Hilton, T.L. An investigation of item bias. Educational and Psychological Measurement, 1968, 28, 61-75.

Cureton, E.E. and Tukey, J.W. Smoothing frequency distributions, equating tests, and preparing norms. American Psychologist, 1957, 6, 404 (Abstract).

Green, D.R. and Draper, J.F. Exploratory studies of bias in achievement tests. A paper presented at the Annual Meeting of the American Psychological Association, Honolulu, September, 1972.

Hoepfner, R., Strickland, G., Stangel, G., Jansen, and Patalino, M. CSE Elementary School Test Evaluations, Center for the Study of Evaluation, Los Angeles, 1970.

Sharon, A.T. and Angoff, W.H. The evaluation of differences in test performance of two or more groups. A paper presented at the annual meeting of the American Psychological Association, Montreal, 1973.

Williams, R. The silent mugging of the black community. Psychology Today. May 1974, 7(12), 32-41, 101.

APPENDIX A

LETTERS TO SUPERINTENDENTS OF THE DISTRICTS,
CHIEF SCHOOL OFFICERS, AND COMMISSIONERS OF THE HEW REGIONS

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION
WASHINGTON, D.C. 20202

Dear Superintendent:

As you probably know, the Emergency School Aid Act (Title VII of Public Law 92-318) provides for grants to local educational agencies (1) to meet special needs incident to the elimination of minority group segregation and discrimination; (2) to encourage the voluntary elimination, reduction, or prevention of minority group isolation; and (3) to aid school children in overcoming the educational disadvantages of minority group isolation.

The U.S. Office of Education is charged with responsibility for evaluating the impact of these grants. In this connection, we will be arranging for a series of special achievement tests in school districts receiving awards. These tests will begin in September, 1973. If your district applies for and receives an award, you may be contacted at a later date regarding the September, 1973 testing.

This letter's purpose is to ask your cooperation in a limited norms testing effort scheduled for May of this year. Minority norms for standardized tests do not exist. We believe that it is important in measuring the achievement of minority group children to do so against norms established for these children themselves, as well as against norms established for the nation as a whole. For this reason, we have drawn a nationally representative sample of 100 minority group isolated schools located in districts which, like yours, meet at least one of the eligibility criteria for ESAA awards. We have arranged for an independent agency, the American College Testing Program, to administer standardized reading and mathematics tests to approximately 30 third graders, 30 fourth graders, and 30 fifth graders in each of these 100 schools, at a time in May to be determined by mutual agreement. Test results will be analyzed to obtain pre-award norms for students in minority group isolated schools.

One or more schools in your district were randomly selected, from the population of minority isolated schools, for inclusion in a national sample. Schools selected in your district are listed at the end of this letter.

We request your permission to have American College Testing Program representatives administer tests to approximately 90 students in each school listed. The tests will require only a few hours of your students' time. The American College Testing Program will conduct the tests and will supply all needed materials. None of your teachers or other staff members will be required to assist during test administration, unless you would prefer to have them present.

Test results for individual students and schools will be completely confidential. A report of norms, summarizing national results for the entire group tested, will be published, but it will not contain any identification of participating schools or districts. We will provide you with copies of this report as soon as it is available.

We are well aware of the fact that the norms testing activity will result in some disruption in your students' scheduling. We ask your cooperation only because we believe that this activity is of considerable importance. We have made every effort to keep our sample small and to reduce the imposition on schools to a bare minimum. In return, we will be able to establish, for the first time, a set of norms for students in minority group isolated schools. These norms will be a valuable tool, not just to us in carrying out our responsibilities for national evaluations, but also to you and to all other educators who are working with minority group children.

We are most anxious to have your cooperation in this effort. Could you appoint a member of your staff who can discuss details with us? It would help considerably if you could telephone us as soon as possible, naming such a point of contact. Your phone call should be to Dr. Michael J. Wargo of my office, at (202) 963-4613.

Correspondence regarding the May, 1973, norms testing activities should be directed to Dr. Michael J. Wargo, Office of Planning Budgeting and Evaluation, Room 4079, U.S. Office of Education, Washington, D.C. 20202. If you wish information on ESAA Grants, however, please direct inquiries to the U.S. Regional Education Office whose address is given at the end of this letter. The participation of your district in these testing activities has no bearing on ESAA grant procedures or decisions.

Sincerely yours,

John W. Evans

John W. Evans
Assistant Commissioner for
Program Planning and Evaluation

88

District Name:

State:

Address of Regional Office for information on ESAA Grants:

School(s) selected in your district:*

---

* If a selected school is missing one or more grades from the three we will be testing (3rd, 4th, and 5th grades), or if it is wholly or partially ungraded at these levels, we will need to use special procedures to select students for testing. We will discuss these with you by telephone.

LETTER TO CHIEF STATE SCHOOL OFFICERS

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION
WASHINGTON, D.C. 20202

The enclosures to this letter are provided to let you know of testing activities we are arranging to carry out in your State, in connection with studies we will be making of Emergency School Aid Act (ESAA) Projects. I am sorry that we have not been able to give you earlier notice of our plans. As you may know, there have been delays in publishing the final versions of the ESAA regulations. These delays have made it necessary for us to omit a number of our originally planned announcement and coordination activities.

In the near future, we will send you a more complete description of our ESAA evaluations. In the mean time, if you have any questions, please call Dr. Michael J. Wargo in my office, phone (202) 963-4613.

Again, I very much regret the delays in our schedule which have prevented earlier notification. I appreciate your consideration and assistance.

Sincerely yours,

John W. Evans
Assistant Commissioner for
Program Planning and Evaluation

Enclosures:

State Districts and Schools Selected
Letter to Superintendents

cc: Coordinator of State Committee on Evaluation and Information Systems

LETTER TO REGIONAL COMMISSIONERS

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION
WASHINGTON, D.C. 20202

The enclosures to this letter are provided to let you know of testing activities we are arranging to carry out in your region, in connection with studies we will be making of Emergency School Aid Act (ESAA) Projects. Similar material is also being sent to Chief State School Officers and to Coordinators of Committees on Evaluation and Information Systems in States affected by these activities.

I am sorry that we have not been able to give you earlier notice of our plans. As you may know, there have been delays in publishing the final versions of the ESAA regulations. These delays have made it necessary for us to omit a number of our originally planned announcement and coordination activities.

In the near future, we will send you a more complete description of our ESAA evaluations. In the mean time, if you have any questions, please call Dr. Michael J. Wargo in my office, phone (202) 963-4613.

Again, I very much regret the delays in our schedule which have prevented earlier notification. I appreciate your consideration and assistance.

Sincerely yours,

John W. Evans

John W. Evans
Assistant Commissioner for
Program Planning and Evaluation

Enclosures:

Regional Districts and Schools Selected
Letter to Superintendents

cc: Regional Senior Program Officer (BEEO)

91

APPENDIX B

RESPONSES TO STUDENT BACKGROUND QUESTIONNAIRE

Form cleared by U.S. Office
of Management and Budget:
OMB No. 51-R-970
Approved through July 31, 1973

District _____

Date _____

Reference
Number _____

STUDENT BACKGROUND QUESTIONNAIRE

For each question, put an "X" in the box or boxes next to the statements that apply to you.

DO NOT ANSWER A QUESTION UNTIL IT HAS BEEN READ ALOUD AND EXPLAINED.

ALL GRADES

1. How long have you been going to this school?

☐	(A) Since preschool or kindergarten.	27%
☐	(B) Since first grade.	24%
☐	(C) Since second grade.	11%
☐	(D) Since third grade.	14%
☐	(E) Since fourth grade.	8%
☐	(F) I am new this year.	15%

2. How many different schools have you gone to since kindergarten?

☐	(A) Only one school.	42%
☐	(B) Two schools.	33%
☐	(C) Three schools.	15%
☐	(D) Four schools.	5%
☐	(E) More than four different schools.	4%

ALL GRADES

3. How many people live with you in your home besides yourself?

☐ (A)	Only one other person.	2%
☐ (B)	Two other people.	7%
☐ (C)	Three or four other people.	29%
☐ (D)	Five to six other people.	32%
☐ (E)	Seven or more other people.	30%

4. Did you do any reading at home during the past two weeks that was not school work?

☐ (A)	Yes.	75%
☐ (B)	No.	25%

5. Did you do any school work at home during the past two weeks?

☐ (A)	Yes.	67%
☐ (B)	No.	32%

6. Is there anyone in your home that can help you with your school work?

☐ (A)	Yes. (ANSWER Q. 7.)	93%
☐ (B)	No. (SKIP Q. 7 AND GO TO Q. 8.)	7%

7. (IF YES TO Q.6) Did you receive any help with your school work at home during the past two weeks?

☐ (A)	Yes.	53%
☐ (B)	No.	41%

ALL GRADES

8. Do the people in your home usually speak another language besides English?

☐ (A)	No, they usually speak English.	74%
☐ (B)	Yes, Spanish.	18%
☐ (C)	Yes, an American Indian language.	1%
☐ (D)	Yes, Chinese.	6%
☐ (E)	Yes, Japanese.	2%
☐ (F)	Yes, some other language.	4%

9. Check the box that best describes yourself.

☐ (A)	Black. (Negro).	63%
☐ (B)	Oriental (Japanese, Chinese, etc.).	1.2%
☐ (C)	American Indian.	3.4%
☐ (D)	White	24%
☐ (E)	Other (Eskimo, Hawaiian, etc.	6%

10. Would you consider yourself of Spanish background (Mexican, Cuban, Puerto Rican, Latin American, etc.)?

☐ (A)	Yes.	21%
☐ (B)	No.	78%

ALL GRADES

11. Which of the following do you have in your home?

☐	(A) Daily newspaper.	65%
☐	(B) Dictionary.	78%
☐	(C) Encyclopedia or other reference books.	59%
☐	(D) Story books.	83%
☐	(E) Magazines.	70%
☐	(F) Record player.	89%
☐	(G) Tape recorder or cassette player.	54%
☐	(H) Color television.	56%
☐	(I) Typewriter.	46%
☐	(J) Automatic dishwasher.	13%
☐	(K) Two or more cars or trucks that run.	47%
☐	(L) Automatic clothes dryer.	31%
☐	(M) A special place to study.	58%

12. How many hours a day do you usually watch television?

☐	(A) Most days, I do not watch television at all.	7%
☐	(B) Most days, I watch television some, but less than one hour.	12%
☐	(C) Most days, I watch television one or two hours.	14%
☐	(D) Most days, I watch television two or three hours.	13%
☐	(E) Most days, I watch television more than three hours.	50%

Breakdown of Item Responses for Student Background Questionnaire by Grade Level

Item	Response	Grade 3	Grade 4	Grade 5
1	A	32%	26%	22%
	B	29%	21%	20%
	C	13%	12%	8%
	D	14%	16%	11%
	E	0%	9%	14%
	F	10%	12%	24%
2	A	47%	42%	37%
	B	32%	33%	32%
	C	12%	14%	18%
	D	4%	5%	7%
	E	3%	4%	5%
3	A	2%	2%	2%
	B	7%	7%	7%
	C	28%	29%	29%
	D	32%	32%	32%
	E	30%	31%	30%
4	A	70%	76%	78%
	B	29%	24%	22%
5	A	63%	67%	71%
	B	36%	32%	28%
6	A	92%	94%	92%
	B	7%	6%	7%
7	A	59%	56%	44%
	B	35%	39%	48%
8	% answering yes			
	A	73%	73%	77%
	B	19%	20%	16%
	C	2%	2%	1%
	D	1%	1%	0.4%
	E	0.3%	0.2%	0.1%
	F	4%	4%	4%

Breakdown of Item Responses for Student Background Questionnaire by Grade Level (continued)

Item	Response	Grade 3	Grade 4	Grade 5
9	A	62%	63%	65%
	B	1.6%	0.9%	0.9%
	C	2.6%	3%	4.5%
	D	25%	24%	23%
	E	6.6%	6.9%	4.3%
10	A	21%	20%	20%
	B	77%	78%	78%
11	% answering yes			
	A	62%	64%	69%
	B	73%	78%	83%
	C	53%	60%	64%
	D	82%	83%	84%
	E	68%	71%	73%
	F	85%	89%	92%
	G	50%	54%	58%
	H	57%	56%	53%
	I	46%	46%	47%
	J	15%	14%	11%
	K	48%	46%	48%
	L	32%	31%	29%
	M	64%	55%	54%
12	A	9%	7%	5%
	B	15%	12%	11%
	C	13%	13%	15%
	D	10%	13%	15%
	E	46%	52%	52%

APPENDIX C

RESPONSES TO STUDENT BACKGROUND QUESTIONNAIRE BY ETHNIC GROUP

Breakdown of Item Responses for Student Background by the Mofied Ethnic Categories

All differences in response patterns between races across all grades are significant beyond the .01 level except for question #7 and part G (tape recorder) on #11.

Item	Response	Black	White	Spanish	Other
1	A	28%	19%	30%	31%
	B	26%	20%	23%	17%
	C	11%	12%	11%	8%
	D	13%	17%	14%	16%
	E	8%	8%	7%	8%
	F	14%	24%	16%	20%
2	A	46%	32%	41%	42%
	B	33%	34%	35%	26%
	C	14%	21%	14%	18%
	D	4%	7%	6%	9%
	E	3%	6%	4%	6%
3	A	2%	1%	2%	3%
	B	6%	7%	7%	9%
	C	25%	44%	29%	27%
	D	34%	31%	31%	33%
	E	34%	17%	30%	29%
4	A	75%	78%	72%	77%
	B	25%	22%	28%	23%
5	A	66%	76%	66%	71%
	B	34%	24%	34%	29%
6	A	95%	94%	88%	94%
	B	5%	6%	12%	6%
7	A	57%	57%	55%	53%
	B	43%	43%	45%	47%
8	% answering yes				
	A	91%	79%	25%	56%
	B	3%	12%	68%	19%
	C	1.2%	0.2%	1.4%	5.6%
	D	0.3%	0.2%	0.4%	8.4%
	E	0.1%	0.1%	0.3%	2%
	F	3.2%	6.3%	2.9%	8%

Breakdown of Item Responses for Student Background by the Modified Ethnic Categories (continued)

Item	Response	Black	White	Spanish	Other
9	A	100%	0%	18%	0%
	B	0%	0%	2%	26%
	C	0%	0%	9%	46%
	D	0%	100%	46%	0%
	E	0%	0%	25%	28%
10	A	0%	0%	100%	0%
	B	100%	100%	0%	100%
11	A	66%	68%	59%	60%
	B	79%	84%	69%	80%
	C	59%	69%	51%	59%
	D	85%	88%	75%	80%
	E	70%	79%	65%	68%
	F	91%	87%	84%	83%
	G	54%	56%	50%	56%
	H	55%	64%	50%	54%
	I	46%	51%	41%	51%
	J	10%	23%	14%	14%
	K	44%	61%	46%	44%
	L	28%	46%	27%	32%
	M	59%	59%	49%	67%
12	% answering yes				
	A	8%	6%	7%	9%
	B	13%	11%	12%	13%
	C	12%	20%	16%	18%
	D	12%	17%	14%	14%
	E	55%	46%	51%	47%

APPENDIX D

ITEMS IDENTIFIED AS POTENTIALLY BIASED, WITH REVIEWERS' COMMENTS

CAT LEVEL 2, FORM A

READING VOCABULARY: Each item consists of a short stimulus phrase, in which one word is printed in bold type, and four single-word response choices. The student is asked to mark the response choice which has the "best" meaning for the word in bold type.

ITEM 38*

American Indian Teacher: Disadvantaged inner-city children understand the word "trip" in a jargon of the drug culture that gives "trip" a different meaning.

Metropolitan Black Administrator and Asian American Teacher: Inner-city poor students, regardless of ethnicity, take very few trips. Those that are taken are never referred to as journeys.

Metropolitan Puerto Rican Teacher: Puerto Rican children would have little experience with trips or other terms referring to trips, such as "journey."

Southwestern Mexican American Teacher: Mexican American children would have little experience with trips or other terms referring to trips.

Southern Black Administrator: "Tractor," one of the distractors, would be completely unfamiliar to inner-city kids and as such may be particularly attractive to inner-city respondents.

ITEM 39

Metropolitan Black Administrator, Metropolitan Puerto Rican Teacher, and Southwestern Mexican American Teacher: Street signs in the ghetto are abbreviated, that is, "Central Ave." A ghetto child relying on his experience may not be able to associate "Ave." with "Avenue."

American Indian Teacher: Children living in rural areas would not be familiar with the term "Avenue."

Southern Black Teacher: "Avenue" and "arena" would have no meaning to a child from a southern town.

Southwestern Mexican American Test Developer: The choices would have little meaning for a rural child.

Southern Black Test Developer: Rural Blacks will be unfamiliar with streets and avenues.

---

* Items in the Level 2 test booklets are not numbered. Item numbers here refer to the sequence number of the item within its subtest.

ITEM 40

Metropolitan Puerto Rican Teacher: "Full" for Spanish American children is something filled, not something added up.

American Indian Teacher: Concepts for total might be "whole" or "all," but not "full." "Full" for some Indian children only relates to objects that are associated with containers.

CAT LEVEL 3, FORM A

READING VOCABULARY: Each item consists of a short stimulus phrase, in which one word is printed in bold type, and four single-word response choices. The student is asked to mark the response choice which has the "best" meaning for the word in bold type.

ITEM 12

Southern Black Administrator: "Pluck" is not a common word in disadvantaged cultural groups. "Pluck" is not part of vocabulary of Southern Blacks.

Metropolitan Black Administrator: Black students do not have the opportunity to pluck strings.

Asian American Teacher: In order to know what is meant, a child must have been exposed to experiences in "plucking" the strings of a violin or other instrument.

Southwestern Mexican American Teacher: "Plucking of strings" is not part of experience of most Spanish American children.

Metropolitan Black Teacher: Some children have experienced "pluck" meaning "to pull" as in "plucking the feathers" from a chicken. They would not realize that "pluck" could also mean "pick."

ITEM 16

Southwestern Mexican American Test Developer: Mexican American children may literally translate "offer" as "ofrecer" which does not necessarily mean "to present."

Southern Black Administrator: Many inner-city and rural Blacks use the term "give a gift" rather than "offer a gift."

ITEM 17

Southwestern Mexican American Teacher, Southern Black Administrator, Southwestern Mexican American Community Leader: The word "ship" is biased in favor of certain regions. The whole concept of vibrating ship would be unfamiliar to many children.

Metropolitan Black Administrator: Most poor kids would not have had the opportunity to experience the sensation of vibrations.

Asian American Teacher: If the word "vibrating" were used with another noun such as "car" which is more common to children of any region and economic level, more children could discern the meaning of "vibrating."

American Indian Teacher: Some children have never seen a ship or even a large body of water.

Metropolitan Black Teacher: The distractor "whirling" may be unfamiliar to disadvantaged children.

Southwestern Mexican American Test Developer: "Vibrating" is a much more familiar word to higher socioeconomic children than to disadvantaged children.

ITEM 23

Southwestern Mexican American Test Developer: Biased against rural children. "Plant" is more closely associated with flowers. Also, Spanish American children are more familiar with tortilla factories.

Puerto Rican Teacher: The word "plant" would mean trees or flowers to a Spanish child and not a factory.

American Indian Teacher: Factories are not a familiar sight in some areas. Some children would think "plant" only applies to trees and flowers and would not associate it with factory.

Asian American Teacher and Metropolitan Black Administrator: This question assumes that the child has been exposed to some concept of industry and he is familiar with the use of the word "plant" in connection with industry. Rural children would not be aware of this usage.

ITEM 29

Southern Black Administrator, Southern Black Teacher, Metropolitan Black Administrator, Asian American Teacher, Southwestern Mexican American Teacher: Building a house that requires a plan is not a common experience of the poor.

Southern Black Administrator: Design would be an unfamiliar concept of the rural child of low income.

Mexican American Community Leader and American Indian Teacher: The distractors cause bias in the use of two similar words, "describe" and "plan." A plan is in part a description.

ITEM 32

Puerto Rican Teacher: Disadvantaged children would associate "principal" with school principal.

American Indian Teacher and Mexican American Teacher: The term "school" as one of the distractors would mislead disadvantaged children in that they only associate "principal" with principal of a school.

Asian American Teacher: Poor inner-city children may correlate the school and the principal as authority with the "law" and may be handicapped by their limited experience.

Metropolitan Black Administrator: The word "school" biases the item for all kids who are familiar with "principal of school."

Southern Black Administrator: "Law," for the children of deprived areas, has a different meaning from what it has for middle-class children.

Southwestern Mexican American Test Developer" Disadvantaged children will associate "principal" with school principal and with disciplinary action.

ITEM 39

Southwestern Mexican American Test Developer and Southwestern Mexican American Teacher: "Deserted" is translated in Spanish as "leaving or breaking" a friendship, which is "discouraging" and "dismaying."

Metropolitan Black Teacher: The first three distractors are all fairly close in that you can be discouraged by all three.

Asian American Teacher: "Discourage" is related to "deserted" in highly transient population.

ITEM 40

Metropolitan Puerto Rican Teacher, Southwestern Mexican American Teacher, and Southwestern Mexican American Test Developer: Children are more familiar with roots of a plant than with roots of problems.

Asian American Teacher: Problems may trigger emotional responses and may be linked with fear, particularly for children who have experienced discrimination or prejudice.

READING COMPREHENSION: Each item requires the student to read a passage and answer several questions measuring comprehension of the content of that passage. Items 11 and 13 refer to a passage describing the geography and resources of Canada. Items 19 and 21 refer to a passage describing the process of erosion, especially as demonstrated by rivers. Item 34 refers to a passage describing the characteristics and the study of the chimpanzee in his natural habitat.

ITEM 11

Southwestern Mexican American Test Developers: Rural children and children who do not live close to oceans are disadvantaged in that they are not familiar with "harbor," "rapids," "ice bound."

Southern Black Administrator and Mexican American Community Leader: The inner-city child has little understanding of the relationship of mining to natural resources.

Metropolitan Black Administrator, Asian American Teacher, and Southwestern Mexican American Teacher: Most ghetto kids have very little experience outside their own community. Their knowledge can be more accurately measured by using a topic with which they are familiar rather than mining and natural resources.

American Indian Teacher: Mining is a term that is associated only with coal in certain parts of the southwest. Children with a second language will probably not comprehend or interpret the mining references in the paragraph.

ITEM 13

Southwestern Mexican American Teacher and Southwestern Mexican American Test Developer: "Land-locked" is an unfamiliar concept, and "unsettled" may be interpreted as "pioneer land" in the historical sense.

American Indian Teacher: Children living in interior regions will have no concept of "land-locked" areas.

Asian American Teacher: Low socioeconomic groups or those who live in the desert or in densely populated areas would have no familiarity of relationships of land-use patterns to natural resources.

Metropolitan Black Administrator: Southerners and other kids with little familiarity of Canada would be unable to relate to the entire question.

ITEM 19

Southwestern Mexican American Test Developers: For some childrer "course" is more closely associated with class or golf course than with area.

Metropolitan Black Administrator: Urban kids, especially those from Southern California, have little experience of this kind.

American Indian Teacher: Some areas of the country are not familiar with rivers; they are more familiar with streams.

Mexican American Teacher: Biased against parts of the country where rivers are an unknown phenomenon.

ITEM 21

Metropolitan Puerto Rican Teacher, Southwestern Mexican American Teacher, and American Indian Teacher: Children in some parts of the country have never seen the places that are described.

Asian American Teacher, Metropolitan Black Administrator, Southern Black Administrator, and Southern Black Administrator: In some regions, children will have this information on which to draw versus others who must entirely deduce the information from the passage.

ITEM 34

Metropolitan Black Administrator: The entire article is biased because of the vocabulary used which is not part of the ghetto experience. Such words as "equatorial," "captivity," "nomadic," "dainty morsels," "vegetarian," and "encroaching" would be unfamiliar.

Southern Black Administrator: Many of the things described will be unfamiliar to low socioeconomic groups.

Southwestern Mexican American Test Developer and Metropolitan Black Teacher: The choices assume that the child is familiar with the behavior of antelopes which may not be true of lower socioeconomic children.

APPENDIX E

SCORE-TO-PERCENTILE-RANK CONVERSION TABLES AND SCHOOL-MEANS-TO-PERCENTILE-RANK CONVERSION TABLES FOR GRADES 3, 4, AND 5

Instruction for using the ESAA-Eligible Minority-Isolated Norm Tables

The tables labeled "Raw Score to Percentile Rank" and "School Means to Percentile Rank" for grades 3, 4, and 5 provide norm information for the individual student level and the school level respectively. The tables are based on the cumulative distribution of the raw achievement scores on the various subtests at the individual student level and the cumulative distribution of the school means at the school level. For both types of tables, percentile ranks are indicated in relation to the raw scores on the original achievement subtests and the debiased achievement subtests. A percentile rank gives the percentage of students in a given reference group that obtained scores equal to or less than a certain score. Percentile ranks represent the relative quality or rank order of each score in comparison with all other scores earned by that reference group, and are comparable from test to test for the same reference group.

If you desire norms (percentile ranks) for an individual student in relation to the other students in the sample, use the "Raw Score to Percentile Rank" tables for the appropriate grade level. Find the raw score under the column for the specific subtest or total of the original or debiased version. Read across the page to the left or right margin on the same line to find the corresponding percentile rank.

The school mean norms are most useful for evaluative purposes, and are presented in the "School Means to Percentile Rank" tables for grades 3, 4, and 5. These norms should be used for comparing the average performance of students at one school relative to the performance of other ESAA-eligible minority-isolated schools across the nation. To use these tables, find the raw score school mean under the column for the specific subtest or total of the original or debiased version. Read across the page to the left or right margin on the same line to find the corresponding percentile rank.

RAW SCORE TO PERCENTILE RANK

GRADE 3

CAT Level 2

%ILE RANK	READING			MATHEMATICS			DEBIASED READING		TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	TOTAL			
99		44-45	85-83		30	101-102		80-82	182-187	180-184	99
98	40	43	82	72				79	180-181	178-179	98
97		42	81		29	100		78	179	176-177	97
96			80			99	37		177-178	175	96
95		41	79	71		98		77	176	173-174	95
94	39		78		28	97		76	173-175	171-172	94
93		40							172	170	93
92			77			96		75	170-171	168-169	92
91		39	76	70				74	168-169	167	91
90	38				27	95			167	165-166	90
89		38	75				36	73	166	164	89
88			74	69		94		72	165	163	88
87		37	73						164	162	87
86					26	93		71	163	161	86
85		36	72	68					161-162	160	85
84						92		70	160	158-159	84
83			71						159	157	83
82	37	35			25	91		69	158	156	82
81				67			35		156-157	155	81
80			70			90		68	155	154	80
79		34							154	153	79
78			69	66		89		67	153	152	78
77	36		68						152	151	77
76		33			24	88		66	151	150	76
75			67	65					150	149	75
74		32	66			87		65	149	148	74
73							34	64	148	147	73
72			65	64		86				146	72
71	35	31	64		23			63	147	145	71
70						85			145-146	144	70

RAW SCORE TO PERCENTILE RANK

GRADE 3

%ILE RANK	READING			MATHEMATICS			DEBIASED READING		TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	TOTAL			
69		30	63	63				62	144	143	69
68						84		61	143	142	68
67			62	62		83	33		142	141	67
66	34	29	61					60	141	140	66
65				61	22	82			140	138-139	65
64		28	60			81		59	138-139	137	64
63			59	60				58	137	136	63
62		27				80	32		136	135	62
61	33		58	59		79		57	134-135	133-134	61
60		26						56	133	132	60
59			57	58	21	78			132	131	59
58		25	56			77		55	131	130	58
57	32		55	57			31	54	130	129	57
56		24				76			129	127-128	56
55			54	56				53	127-128	126	55
54	31	23	53			75		52	126	125	54
53				55	20	74	30		125	124	53
52			52			73		51	124	123	52
51		22	51	54				50	123	122	51
50	30					72	29		121-122	121	50
49		21	50	53		71		49	120	119-120	49
48			49		19			48	119	118	48
47	29	20	48	52		70			117-118	116-117	47
46						69	28	47	116	115	46
45		19	47	51				46	114-115	114	45
44	28		46		18	68		45	113	112-113	44
43		18		50		67	27		112	111	43
42			45	49		66		44	111	110	42
41	27		44				26		109-110	109	41
40		17		48		65		43	108	108	40

RAW SCORE TO PERCENTILE RANK

GRADE 3

%ILE RANK	READING			MATHEMATICS			DEBIASED READING		TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	TOTAL			
39			43		17	64		42		107	39
38	26			47		63	25		107	106	38
37		16	42	46		62		41	106	105	37
36	25			45		61			104-105	104	36
35			41				24	40	103	102-103	35
34		15		44	16	60			102	101	34
33	24		40			59		39	101	100	33
32			39	43			23	38	100	99	32
31				42		58			99	98	31
30	23	14	38		15	57		37	97-98	96-97	30
29			37	41		56	22		96	95	29
28								36	95	94	28
27	22		36	40		55			94	93	27
26				39		54	21	35	92-93	91-92	26
25		13	35		14				91	90	25
24	21			38		53		34	90	89	24
23			34	37		52	20		88-89	88	23
22			33	36		51		33	87	86-87	22
21	20	12		35	13	50	19	32	86	85	21
20			32			49			85	84	20
19	19			34		48		31	83-84	83	19
18			31	33	12	47	18		82	82	18
17				32		46		30	81	80-81	17
16	18	11	30			45			79-80	78-79	16
15				31	11	44	17	29	77-78	77	15
14	17		29	30		43			76	75-76	14
13				29		42	16	28	74-75	73-74	13
12		10	28	28		40-41		27	73-72	72	12
11	16		27	27	10	39			71	70-71	11
10				26		38	15	26	68-70	68-69	10

RAW SCORE TO PERCENTILE RANK

GRADE 3

%ILE RANK	READING			MATHEMATICS			DEBIASED READING		TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	TOTAL			
9	15	9	26	25		37		25	66-67	66-67	9
8				24	9	35-36	14		64-65	63-65	8
7	14		25	23		34		24	62-63	61-62	7
6		8	24	21-22	8	32-33	13		59-61	59-60	6
5	13		23	19-20		29-31	12	23	56-58	55-58	5
4	12	7		17-18	7	27-28		22	52-55	52-54	4
3	11		21-22	14-16		23-26	11	21	48-51	47-51	3
2	10	6	20	12-13	5-6	20-22	10	19-20	43-47	42-46	2
1	0-9	0-5	0-19	0-11	0-4	0-19	0-9	0-18	0-42	0-41	1

RAW SCORE TO PERCENTILE RANK

GRADE 4

CAT Level 3

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
99	34-40	35-42	67-82	45-48	25	64-73	29-32	32-37	59-69	124-155	116-142	99
98	32-33	32-34	62-66	43-44	22-24	61-63	28	29-31	55-58	117-123	111-115	98
97	30-31	30-31	58-61	41-42	21	58-60	26-27	28	52-54	110-116	105-110	97
96	29	29	55-57	40	20	57	25	27	50-51	106-109	102-104	96
95	28	28	54	39	19	55-56		26	48-49	104-105	99-101	95
94	27	27	52-53	38		54	24	25	47	102-103	97-98	94
93	26	26	50-51	37	18	53	23	24	46	100-101	95-96	93
92			49	36		52			44-45	99-98	93-94	92
91	25	25	48	35		51	22	23		96-97	92	91
90		24	47		17	50			43	94-95	90-91	90
89	24		46	34			21	22	43	92-93	88-89	89
88	23	23	45			49			41	91	87	88
87			44	33		48		21	40	89-90	85-86	87
86					16	47	20		39	88	84	86
85	22	22	43	32		46				87	83	85
84			42					20	38	86	82	84
83						45	19			84-85	81	83
82	21	21	41	31	15				37	83	79-80	82
81						44				82	78	81
80			40	30			18	19	36	81	77	80
79	20		39			43				80	76	79
78		20		29					35	79	75	78
77			38		14	42				78	74	77
76						41	17	18	34	77		76
75	19		37	28						76	73	75
74		19	36			40			33	75	72	74
73				27				17		74	71	73
72					13	39	16		32	73	70	72
71	18	18	35							72	69	71
70				26		38			31	71	68	70

RAW SCORE TO PERCENTILE RANK

GRADE 4

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
69			34			37		16		70	67	69
68									30	69	66	68
67	17	17	33	25	12	36	15			68	65	67
66											64	66
65						35			29	67	63	65
64			32	24				15		66	62	64
63	16					34				65		63
62		16					14		28	64	61	62
61			31	23	11					63	60	61
60						33			27		59	60
59								14		62		59
58	15		30	22		32	13			61	58	58
57		15							26	60	57	57
56			29			31					56	56
55					10					59		55
54				21					25	58	55	54
53	14		28			30	12	13				53
52										57	54	52
51		14		20		29				56	53	51
50			27						24			50
49					9					55	52	49
48						28						48
47	13		26	19			11		23	54	51	47
46						27		12			50	46
45		13								53		45
44										52	49	44
43			25	18		26			22			43
42					8					51	48	42
41	12						10			50		41
40			24			25		11			47	40

RAW SCORE TO PERCENTILE RANK

GRADE 4

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCEPT	TOTAL	VOCAB	COMPR	TOTAL			
39				17					21	49	46	39
38		12										38
37						24				48	45	37
36	11		23									36
35				16		23			20	47	44	35
34					7		9					34
33						22				46	43	33
32			22					10				32
31		11		15					19	45		31
30	10										42	30
29						21				44		29
28			21								41	28
27				14			8		18	43		27
26					6	20					40	26
25		10	20					9		42	39	25
24	9									41		24
23				13		19			17		38	23
22										40	37	22
21			19				7			39		21
20				12		18			16		36	20
19	8	9	18		5			8		38		19
18						17				37	35	18
17											34	17
16				11					15	36		16
15			17			16	6			35	33	15
14	7	8						7		34	32	14
13				10	4				14		31	13
12			16			15				33	30	12
11										32		11
10	6	7	15	9		14	5		13	31	29	10

RAW SCORE TO PERCENTILE RANK

GRADE 4

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
9					3	13		6		30	28	9
8			14	8					12	29	27	8
7	5		13			12				27-28	26	7
6				7	2	11	4	5	11	26	24-25	6
5		5	12			10			10	24-25	22-23	5
4	4		11	6		9	3	4	9	22-23	20-21	4
3		4	9-10	5	1	8			8	19-21	18-19	3
2	3		8	4		7	2	3	7	16-18	15-17	2
1	0-2	0-3	0-7	0-3	0	0-6	0-1	0-2	0-6	0-15	0-16	1

RAW SCORE TO PERCENTILE RANK

GRADE 5

CAT Level 3

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
99	38-40	38-42	75-82	47-48	24-25	69-73	32	34-37	65-69	138-155	129-142	99
98	37	37	72-74	46	23	66-68	31	33	63-64	133-137	124-128	98
97	36	35-36	69-71	45		65	30	32	60-62	129-132	121-123	97
96	35	33-34	66-68		22	64		31	58-59	125-128	118-120	96
95	34	32	64-65	44	21	63	29	30	57	123-124	116-117	95
94	33		62-63			62	28	29	55-56	121-122	114-115	94
93	32	31	61	43				28	54	119-120	113	93
92	31	30	60		20	61	27			117-118	111-112	92
91			58-59	42		60		27		115-116	109-110	91
90	30	29	57						52	113-114	108	90
89		28	56	41	19	59	26	26	51	112	107	89
88	29					58			50	111	105-106	88
87		27	55				25		49	109	104	87
86	28		54	40		57		25		107-108	102-103	86
85			53						48	106	101	85
84	27	26		39		56	24		47	105	100	84
83			52		18			24		104	99	83
82			51			55			46	103	98	82
81	26	25	50	38			23			102		81
80			49			54		23	45	101	97	80
79										100	96	79
78	25	24	48	37		53	22		44	99	94-95	78
77					17					98		77
76			47					22	43	97	93	76
75	24		46	36		52			42	96	92	75
74		23					21			95	91	74
73			45			51			41	94		73
72	23			35				21		93	90	72
71			44			50			40		89	71
70		22			16		20			92	88	70

RAW SCORE TO PERCENTILE RANK

GRADE 5

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
69			43	34		49		20	39	91	87	69
68	22									90	86	68
67			42			48					85	67
66		21		33					38	89		66
65			41				19			88	84	65
64	21					47		19	37	87	83	64
63			40							86	82	63
62		20		32	15	46			36	85		62
61							18				81	61
60	20		39			45		18		84	80	60
59				31					35	83		59
58		19	38							82	79	58
57						44	17		34		78	57
56	19		37	30	14					81	77	56
55						43		17		80	76	55
54									33	79		54
53		18	36	29						78	75	53
52	18					42	16		32	77	74	52
51										76	73	51
50			35	28		41					72	50
49					13			16	31	75	71	49
48	17	17	34			40	15			74		48
47				27						73	70	47
46						39			30		69	46
45			33							72	68	45
44				26		38		15	29	71	67	44
43	16		32		12		14			70		43
42		16				37				69	66	42
41				25					28		65	41
40			31			36				68		40

RAW SCORE TO PERCENTILE RANK

GRADE 5

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
39	15							14		67	64	39
38				24		35	13		27	66	63	38
37		15	30		11					65	62	37
36				23						64	61	36
35			29			34			26		60	35
34										63		34
33	14			22		33	12	13		62	59	33
32			28		10				25	61	58	32
31		14		21		32					57	31
30										60		30
29			27			31			24		56	29
28	13			20						59	55	28
27					9	30	11			58	54	27
26		13	26			29			23	57		26
25				19						56	53	25
24						28				55	52	24
23	12		25					11	22	54	51	23
22				18	8	27	10			53	50	22
21		12										21
20			24			26			21	52	49	20
19	11			17		25				51	48	19
18							9			50	47	18
17			23		7	24		10	20	49	46	17
16		11		16						48	45	16
15	10		22			23			19	47	44	15
14				15		22	8			46	43	14
13					6			9		45	42	13
12	9	10	21	14		21			18	44	41	12
11										43	40	11
10			20			20	7		17	42	39	10

RAW SCORE TO PERCENTILE RANK

GRADE 5

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
9	8		19	13	5	19		8		41	38	9
8		9				18			16	39-40	37	8
7			18	12			6			38	35-36	7
6	7		17	11	4	17		7	15	36-37	34	6
5		8				16			14	34-35	32-33	5
4	6		16	10	3	14-15	5			33	30-31	4
3		7	14-15	9	2	13		6	13	31-32	28-29	3
2	5	6	13	8	1	11-12	4	5	11-12	27-30	25-27	2
1	0-4	0-5	1-12	0-7	0	0-10	0-3	0-4	1-10	0-26	0-24	1

SCHOOL MEANS TO PERCENTILE RANK

GRADE 3

CAT Level 2

%ILE RANK	READING			MATHEMATICS			DEBIASED READING		TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	TOTAL			
99		36	71	66	25	92		69	163	162	99
98	35				24	88					98
97			68				33	66	150	149	97
96		32	66-67	63		85-86		65	149	148	96
95	34			62	23						95
94		30		61					148	147	94
93				60						146	93
92				59		81		61			92
91	33				22	80		60	142	139-140	91
90		28	61				31		141		90
89				58		79			140	138	89
88								59			88
87	32					78		58	138	137	87
86		27	59						137	136	86
85								57	136		85
84				57	21	77	30			134-135	84
83			58						135		83
82								56		133	82
81									134		81
80	31		57			76					80
79										132	79
78		26		56				55	133		78
77					20						77
76			56								76
75									132	130	75
74						75		54	131		74
73				55			29				73
72		25	55						130	129	72
71	30							53			71
70								53	129	128	70

SCHOOL MEANS TO PERCENTILE RANK

GRADE 3

%ILE RANK	READING			MATHEMATICS			DEBIASED READING		TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	TOTAL			
69				54					129	127	69
68			54			74			127		68
67								52	125		67
66		24			19					124	66
65			53						124	123	65
64						73					64
63				53					123	122	63
62			52				28	51			62
61	29					72				121	61
60									122		60
59		23									59
58			51			71				120	58
57				52				50	121		57
56										119	56
55						70			120		55
54							27				54
53				51				49			53
52	28		50						118	117	52
51		22				69					51
50				50	18				117		50
49						68				116	49
48											48
47				49				48	116		47
46			49			67					46
45							26			115	45
44				48		66			115		44
43	27									114	43
42		21									42
41						65		47	114	113	41
40			48	47							40

SCHOOL MEANS TO PERCENTILE RANK

GRADE 3

%ILE RANK	READING			MATHEMATICS			DEBIASED READING		TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	TOTAL			
39									113	112	39
38										111	38
37						64		46	112		37
36			47	46			25		111	110	36
35	26								110		35
34		20				63					34
33								45	109	108	33
32			46	45	17						32
31						62					31
30									107		30
29		19	45					44		106	29
28											28
27	25			44		61	24		106		27
26			44					43			26
25		18							105	104	25
24											24
23			43			60		42			23
22											22
21				43	16	59			103	102	21
20	24	17	42					41			20
19						58			102	101	19
18											18
17				42		57			101	100	17
16											16
15			41		15			40	100	99	15
14									99		14
13	23	16		41		56				98	13
12			40				22				12
11								39	98	97	11
10				40							10

SCHOOL MEANS TO PERCENTILE RANK

GRADE 3

%ILE RANK	READING			MATHEMATICS			DEBIASED READING		TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	TOTAL			
9					14	55			97	96	9
8	22		39	39				38	96	95	8
7		15									7
6				37		54	21		93	92	6
5			38					37	92	91	5
4				36	13						4
3		14		35		48			88	87	3
2	19	13	37	34	12		19	36	87	86	2
1	15	11	36	23	10-11	39	14	35	77	76	1

SCHOOL MEANS TO PERCENTILE RANK

GRADE 4

CAT Level 3

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
99				34		48				90		99
98		22	42	33	15	47		20		89		98
97	21	21		32	14		18		37	88	85	97
96								19				96
95			41			43				83	78	95
94		20		29		42	17		36	80	77	94
93			40		13	41		18		79		93
92	19		39	28					35		74	92
91						40				78	73	91
90		19		27			16			77		90
89			37			39				76		89
88	18							17	33			88
87			36	26						75	71	87
86						38			32	74	70	86
85		18			12							85
84			35	25			15					84
83						37				72	68	83
82	17								31			82
81				24							67	81
80						36		16				80
79			34							70		79
78									30		66	78
77		17				35						77
76				23	11		14			69		76
75	16										65	75
74			33									74
73						34			29		64	73
72								15		67	63	72
71						33				66		71
70			32							65	62	70

RAW SCORE TO PERCENTILE RANK

GRADE 4

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
69		16		22						64		69
68									28		60	68
67	15					32	13			63		67
66			31								59	66
65										62		65
64					10			14				64
63									27	61	58	63
62			30									62
61		15									57	61
60										60		60
59				21		31						59
58	14								26			58
57			29									57
56										59	56	56
55												55
54								13				54
53									25			53
52			28									52
51		14			9	30				58		51
50											55	50
49												49
48	13			20					24			48
47												47
46												46
45										57	54	45
44						29	11					44
43									23			43
42			26					12				42
41										56		41
40				19							53	40

SCHOOL MEANS TO PERCENTILE RANK

GRADE 4

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
39		13				28						39
38											52	38
37										54	51	37
36					8							36
35	12		25			27			22	53	50	35
34												34
33										52		33
32				18		26					49	32
31										51		31
30											48	30
29						25		11				29
28							10			49		28
27		12	24								46	27
26				17					21			26
25						24				48		25
24											45	24
23												23
22												22
21				16	7	23						21
20	11									47	44	20
19		23										19
18												18
17				15		22			20			17
16								10		46	43	16
15		11										15
14						21						14
13				14								13
12			22				9					12
11					6	20						11
10									19		40	10

SCHOOL MEANS TO PERCENTILE RANK

GRADE 4

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
9										42	39	9
8	10			13						41		8
7		10	21			19					38	7
6					5			9				6
5				12	4	18			18	40	37	5
4										39		4
3					3	16-17	8				36	3
2	9		19					8	17			2
1	8	7-8	18	8-9	1	10-14	7	6	16	29-33	27-31	1

SCHOOL MEANS TO PERCENTILE RANK

CAT Level 3

GRADE 5

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
99	27		53	39	18	57			46	107	103	99
98	26	26					23	24				98
97		25	50			53	22	23	45	101	97	97
96			49	37						99	95	96
95	24					52	21			98		95
94			48						43		93	94
93		24	47		16			22	43	97	92	93
92	23			35		51	20					92
91			46						41		91	91
90		23				50		21		95	90	90
89						49				94	89	89
88			45						40			88
87	22			32	15	48	19			92	88	87
86		22						20	39			86
85				31		47				91	87	85
84			43								86	84
83						46			38			83
82										90	85	82
81	21	21	42					19				81
80				30		45			37	89		80
79							18			88		79
78											83	78
77					14					87	82	77
76						44				86		76
75									36			75
74		20	40				20			84	80	74
73	20									83	79	73
72				29		43		18				72
71										82	78	71
70												70

SCHOOL MEANS TO PERCENTILE RANK

GRADE 5

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
69									35	81	77	69
68			39			42						68
67												67
66					13					80	76	66
65		19					19					65
64						41						64
63	19								34	79	75	63
62			38	28								62
61										78		61
60											74	60
59						40		17				59
58										77		58
57			37								73	57
56									33			56
55												55
54		18					18			76	72	54
53					12							53
52												52
51	18		36	27		39				75		51
50											71	50
49												49
48									32	74		48
47												47
46								16			70	46
45												45
44										73		44
43			35								69	43
42		17		26		38						42
41									31			41
40							15			72		40
39	17										68	39
38												38
37					11							37
36			34						30			36
35				25		37					67	35

SCHOOL MEANS TO PERCENTILE RANK

GRADE 5

%ILE RANK	READING			MATHEMATICS			DEBIASED READING			TOTAL BATTERY	DEBIASED TOTAL BATTERY	%ILE RANK
	VOCAB	COMPR	TOTAL	COMPU	CONCPT	TOTAL	VOCAB	COMPR	TOTAL			
34								15		70		34
33											66	33
32			33						29	69		32
31	16	16					14					31
30				24						68		30
29			32						28		64	29
28												28
27										67		27
26			31								63	26
25								14	27	66		25
24				23		34						24
23	15		30				13			65	62	23
22		15			10						61	22
21												21
20						33			26			20
19			29							63	60	19
18				22								18
17								13				17
16											59	16
15	14					32				62		15
14		14	28				12		25			14
13												13
12						31				61	58	12
11				21								11
10					9					60	57	10
9	13		27					12	24			9
8						30					56	8
7		13					11			59		7
6				20							55	6
5	12				8	29			23	58		5
4			26									4
3								11		56	53	3
2		12				27				55		2
1	11	11	25	17-18	1	22-26	10	10	22	53	50-52	1

MEMORANDUM

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

TO : Readers of: Achievement Test Restandardization DATE: January 17, 1975

FROM : Michael J. Wargo, ESAA Evaluation Program Officer

SUBJECT: (1) Ethical principles on release of test items to general public.
(2) Hypothetical examples of biased test items.

Reviewers of Achievement Test Restandardization have suggested that the Office of Education release to the public some examples of achievement test items determined to be biased against minority students. Such a practice would be inconsistant with the Ethical Standards of Psychologists (American Psychological Association) and the Standards for Development and Use of Educational and Psychological Tests (National Council on Measurement in Education, American Educational Research Association, and American Psychological Association) which forbid the release of standardized test items to the general public on the grounds that such release would invalidate test items and possibly the entire test. Qualified test users, to whom the above restriction does not apply, can identify items that were determined to be baised by matching the item numbers in Appendix D with the items on a copy of the published test.

Since this report does not include actual examples of test item bias, the following will provide two hypothetical examples for illustrative purposes. Examples are structured and formatted as were the actual items determined to be biased by the reported study.

Example Biased Items

The following examples are designed to illustrate how an item in a standardized test can be biased against particular subgroups of students. The examples were designed for students in grades 3, 4, & 5 and could be part of a Reading Vocabulary subtest of any standardized achievement test.

Instructions: For each of the items below choose the word with the best meaning for the word underlined. Circle the word with the best meaning.

1. Comfortable den

   . bath

   . animal

   . study

   . sofa

2. fast boulevard

   . traffic

   . street

   . trip

   . stream

Explanation of possible bias

Item 1: The higher the socioeconomic status of a family the greater the likelihood that the home of the family will contain a spare room that might be referred to as a "den." Further, if such a room exists, the use of it as a "study" would tend to increase as the socioeconomic status of the family increases. Therefore, one would expect poor minority group children, living in homes with large families, to have had less exposure to the word "den" than their more advantaged peers and to have a less clear understanding of its possible use as a "study." In short, Item 1 might be biased against such children in its ability to measure reading vocabulary.

Item 2: Children from higher socioeconomic status families generally are exposed to more varied reading materials and travel experiences. The probability of such children being exposed to the word "boulevard" and experiencing the sight of a "boulevard" is greater for them than for their more disadvantaged peers. Inner city and rural minority group members are therefore less likely to be familiar with the word "boulevard" and the item may be biased against them in measuring their reading achievement.